STRUCTURAL AND GENETIC ANALYSIS

OF HEPATITIS G VIRUS / GB VIRUS-C

NARCISA MANUELA CUCEANU

A thesis submitted for the degree of Doctor in Philosophy

The University of Edinburgh

1999

To my parents

DECLARATION

The experiments described in this thesis have been performed by the author unless otherwise stated.

The contents of this thesis have been composed by the author.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Dr. P. Simmonds for his assistance and advice throughout the work carried out in this thesis. I am also grateful to Dr. D.B. Smith for his helpful comments on the manuscript and to Dr. J. Mellor for assistance with the cell culture work.

Many thanks go to all my colleagues from Molecular Virology group for their help and support. Dr. D. Apps (Department of Biochemistry) has kindly allowed the use of the Abbé refractometer for the sucrose density study.

I wish to express my gratitude to Professor Kenneth Murray for his moral support and encouragement throughout my Ph.D. stage in Edinburgh.

My special thanks go to Arjen for his constant love, patience and support and to my family and Arjen's family for encouragement during my work.

This work was supported by a grant from Darwin Trust, Edinburgh and in part by a grant from Die Norken Stiftung, Switzerland.

CONTENTS

CHAPTER 2. MATERIALS AND METHODS

PUBLICATIONS

Smith, D.B., Cuceanu, N., Davidson, F., Jarvis, L.M., Mokili, J.L., Hamid, S., Ludlam, C.A., Simmonds, P. (1997). Discrimination of hepatitis Gvirus/GBV-C geographical variants by analysis of the 5' non-coding region. *Journal of General Virology* 78, 1533-1542.

Smith, D.B., Basarabas, M., Frost, S., Haydon, D., Cuceanu, N., Prescott, L., Kamenka, C., Millband, D., Sathar, M.A., Simmonds, P.(1999). Phylogenetic analysis of GBV-C/hepatitis G virus. *Journal of General Virology* submitted.

ORAL PRESENTATIONS

Cuceanu, N., Jarvis, L., Simmonds, P. Buoyant density of hepatitis G virus. Society for General Microbiology, University of Warwick, March 1996 (presented by P. Simmonds).

Cuceanu, N., Smith, D.B., Davidson, F., Jarvis, L.M., Mokoli,J., Simmonds, P. Phylogenetic analysis of HGV/GBV-C 5 non-coding region sequences. Society for General Microbiology, Heriot Watt University, Edinburgh. March 1997.

POSTER PRESENTATION

Molecular studies of Hepatitis G virus/GB Virus-C at Darwin Trust Symposium, November 1998.

ABSTRACT

Hepatitis G virus (HGV) or GB virus C (GBV-C) is a novel member of the *Flaviviridae* family, discovered by two independent research groups. HGV/GBV-C has a single-stranded, positive-sense RNA genome of approximately 9.4 kb in length which contains a single open reading frame (ORF). The genome organization is similar to that of hepatitis C virus (HCV) with 25% amino acid identity. Extensive studies have been carried out to assess the clinical importance and epidemiology of this new agent.

This thesis describes the genetic analysis of the heterogeneity of HGV/GBV-C and the characterization of the terminal regions of the viral genome. The sequence diversity across the HGV/GBV-C genome was significantly lower than that observed with HCV isolates. Comparative analysis of twenty-seven complete genome HGV/GBV-C sequences indicated the presence of four phylogenetic groups and this study demonstrated that these groupings could be reproduced by analysis of the 5'-untranslated region (5'-UTR) and of various sub-fragments. At the same time, the analysis of the 5'-UTR variability indicated the existence of group-specific polymorphisms, many of which are covariant and consistent with the proposed secondary structure of this region.

An important difference between the polyproteins of HGV/GBV-C and HCV is the absence of a putative HGV/GBV-C core protein which is usually encoded at the 5'-end of the genome of flaviviruses. The buoyant density of

HGV/GBV-C particles in human plasma was estimated to be between 1.07-1.12 g/ml, much lower than that of the other members of *Flaviviridae* family, except HCV. No HGV/GBV-C RNA was detected in fractions with densities higher than 1.17 g/ml, which is expected for virus particles in immune complexes, or in fractions with densities higher than 1.21 g/ml, density range of HCV nucleocapsids. These biophysical properties correlate with the absence of a core-like protein in the genome of HGV/GBV-C isolates from different phylogenetic groups. The absence of the HGV/GBV-C nucleocapsid was also revealed by the sequence analysis data since no conserved open reading frame capable of encoding a core-like protein was identified.

Generally, the untranslated regions at the 5' and 3' termini of a RNA virus genome contain regulatory elements important for viral RNA replication, transcription, translation and viral packaging. A comprehensive comparison and analysis of the primary sequence and secondary structure of the 3'-UTR of different HGV/GBV-C isolates allowed the construction of a common secondary structure model for this region and the identification of structural elements that may be involved in viral replication.

Finally, the susceptibility of various types of cultured cells (peripheral blood mononuclear cells, U937 and HepG2 cells) to HGV/GBV-C infection was examined in an attempt to develop a reliable cell culture system that will allow detailed investigation of the *in vitro* replication of HGV/GBV-C.

Together these investigations clarify basic features of the structure, replication and variability of HGV/GBV-C.

ABBREVIATIONS

ALT	Alanine aminotransferase
AS	antisense primer
bp	base pairs
BVDV	bovine viral diarrhoea virus
BSA	bovine serum albumin
BW	bead wash
cDNA	complementary DNA
CMV	cytomegalovirus
^{0}C	degrees Celsius
DEPC	diethylpyrocarbonate
DNA	deoxyribonucleic acid
dNTPs	deoxyribonucleotides
ds	double stranded
EBV	Epstein Barr virus
ELISA	enzyme linked immunosorbent assay
GBV-A	GB virus A
GBV-B	GB virus B
GBV-C	GB virus C
HAV	hepatitis A virus
HBV	hepatitis B virus

HCV	hepatitis C virus
HCHV	hog cholera virus
HDV	hepatitis D virus
HEV	hepatitis E virus
HFV	hepatitis F virus
HGV	hepatitis G virus
HVR	hypervariable region
I	inner primer
IFN	interferon
Ig	immunoglobulin
IL	interleukin
IRES	internal ribosome entry site
IPTG	isopropyl-thiogalactosidase
IVDU	intravenous drug user
JEV	Japanese encephalitis virus
kDa	kilodalton
kb	kilobase
LiPA	line probe assay
LSH	Long Stable Hairpin
μg	microgram
μl	microlitre
mg	milligram
ml	millilitre

mM	millimolar
NANBH	non-A non-B hepatitis
ng	nanogram
nm	nanometre
O	outer primer
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBMC	peripheral blood mononuclear cell
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PT-NANBH	post-transfusion associated NANBH
RdRp	RNA dependent RNA polymerase
RFLP	restriction fragment length polymorphism
RIBA	recombinant immunoblot assay
RNA	ribonucleic acid
rpm	revolutions per minute
RT	reverse transcription
S	sense primer
SDS	sodium dodecyl sulphate
SE	South-East
ss	single stranded
TE	tris-EDTA buffer
UTR	untranslated region

UV	ultra violet
V	volts
YFV	yellow fever virus

CHAPTER 1

1. GENERAL INTRODUCTION: HEPATITIS G VIRUS/GB VIRUS-C

1.1 VIRAL HEPATITIS AND THE SEARCH FOR NANB AGENTS

Viral hepatitis represents an important public health problem throughout the world. Initially, the disease was known to be caused by two viruses - hepatitis A virus (HAV) and hepatitis B virus (HBV) - each with its own specific mode of transmission and replication. HAV, originally classified as an Enterovirus (type 72), is a small unenveloped RNA virus which has been classified as member of the *Picornaviridae* family and causes infectious and epidemic hepatitis transmissible by faecal-oral route (Melnick, 1982). HBV, a double-stranded DNA virus which replicates by reverse transcription (RT), belongs to the hepadnavirus group. It is endemic in the human population and closely related viruses infect woodchucks, squirrels and ducks (Zuckerman, 1990).

Infections due to these two viruses can be identified by tests for specific serological and enzymatic markers such as: hepatitis B surface antigen (HBsAg), anti-hepatitis B core antibody (anti-HBc), immunoglobulin M antibody to HAV. However, it was reported in the 1970s that a number of cases of transfusion-associated hepatitis were serologically unrelated to either HAV or HBV, and so were described as non-A, non-B hepatitis (NANBH). NANBH appears to be a mild, frequently asymptomatic, disease only causing jaundice in some cases. Acute infection can progress to chronicity in a higher proportion (50%) compared to approximately 5% for HBV infection. At the same time, the

incubation period from exposure to liver function abnormalities is shorter (8 weeks) than usually observed for HBV infection (longer than 12 weeks) but longer than that for HAV infection (between 3 and 5 weeks) (Conn and Atterbury, 1993).

Since cytomegalovirus (CMV) and Epstein-Barr virus (EBV) are known to cause liver damage and these viruses can be transmitted by blood transfusion, it was suggested that these viruses may be responsible for some of NANBH cases (Luby *et al.* 1974). However, serological investigations showed that CMV or EBV are rarely involved, so they cannot be considered as causative agents of NANBH (Feinstone *et al.* 1975; Alter *et al.* 1978). As a result, diagnosis of NANBH was made by exclusion of HAV, HBV and herpesviruses such as CMV and EBV.

The first direct evidence for a third human hepatitis virus was provided by immunofluorescence microscopy in the mid1970s (Rizzetto *et al.* 1977). A new viral antigen, identified in the nuclei of hepatocytes from Italian patients with chronic hepatitis B, proved to be a small single-stranded circular RNA virus with a number of similarities to certain plant viral satellites and viroids. It was named hepatitis D virus (HDV) and it was shown that it requires HBV to establish infection, and that it causes both fulminant hepatitis and accelerated progression of pre-existing hepatitis B virus infection (Polish *et al.* 1993). These particular features indicated that HDV cannot be considered as a potential NANBH agent. At the same time, epidemiological studies revealed that NANBH represents more than 90% of transfusion-associated hepatitis cases in

The United States (Aach *et al.* 1981). Screening of blood donors for serum alanine aminotransferase levels (ALT) (Aach *et al.* 1981) and for antibodies to hepatitis B core antigen (anti-HBc) (Koziol *et al.* 1986) was proposed in order to reduce the high risk of transmitting NANB agent by blood transfusions. Different investigations using these "surrogate markers" for the detection of NANBH virus carriers indicated that exclusion of anti-HBc positive donors would prevent 42-43% of the NANBH cases (Koziol *et al.* 1986; Sugg *et al.* 1988).

The hypothesis that a novel viral hepatitis agent was responsible for NANBH was further supported by epidemiological studies in humans which showed that hepatitis was transmitted by serum derived from patients with chronic and acute hepatitis when there was no serological evidence of type A or B hepatitis (Prince *et al.* 1974; Alter *et al.* 1978). In addition, studies in primates (Feinstone *et al.* 1975; Hoofnagle *et al.* 1977) indicated that the NANBH agent may be a small, enveloped virus, between 30 and 60 nm in diameter, that is transmissible to chimpanzees (He *et al.* 1987) and can be inactivated by chloroform treatment (Feinstone *et al.* 1983). Major progress was subsequently achieved using a recombinant complementary DNA (cDNA) approach in 1989 (Choo *et al.* 1989). A cDNA library derived from infectious plasma containing the NANBH agent was produced using *Escherichia coli* bacteriophage λgt11 and then screened for clones expressing viral antigen using serum from a chronic NANBH patient. This immunoscreening procedure led to the identification and isolation of a positive cDNA clone designated 5-1-1. Hybridization analysis

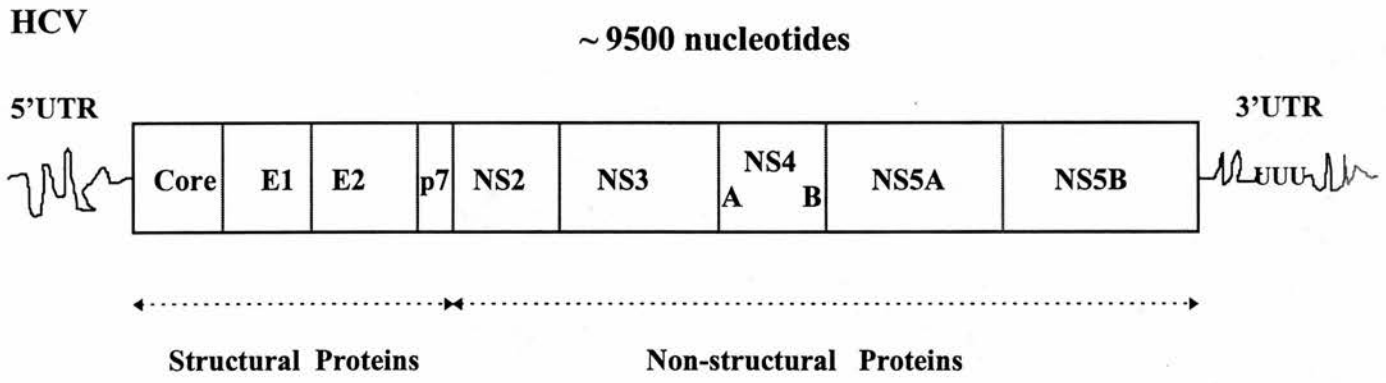
indicated that this clone was not derived from host DNA but from an exogenous RNA molecule associated with NANBH infection. This single-stranded RNA molecule with an approximate size of 9,400 nucleotides was designated hepatitis C virus (HCV) (Choo *et al.* 1989; Choo *et al.* 1991). The HCV viral genome of positive polarity consists of a single continuous open reading frame (ORF), encoding a single polyprotein and flanked by untranslated regions at both its 5' and 3'ends (Figure 1.1). The polyprotein of approximately 3,010 amino acids is cleaved post-translationally into individual viral proteins: core (C), E1, E2 (envelope), p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B (non-structural) proteins (Choo *et al.* 1991; Takamizawa *et al.* 1991).

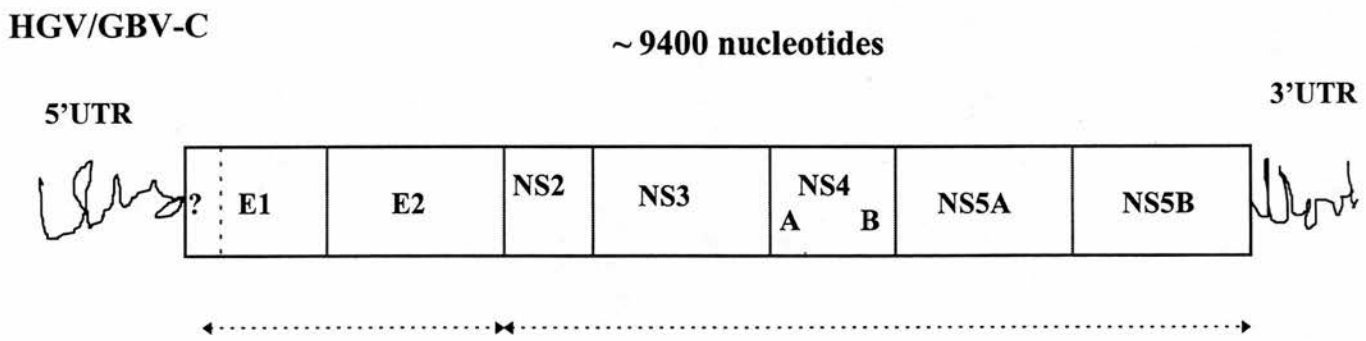
Comparison with other positive-strand RNA virus groups revealed that HCV is distantly related to the pestiviruses and flaviviruses, both of which are members of the *Flaviviridae* family (Kato *et al.* 1990; Miller and Purcell, 1990; Koonin, 1991). For example, sequence alignment of amino acid sequences of the RNA-dependent RNA polymerases (NS5B region) of all positive-strand RNA virus groups identified three large phylogenetic supergroups of RNA polymerases (RdRp); supergroup II included pestiviruses, HCV, and flaviviruses together with carmoviruses, tombusviruses and single-stranded RNA bacteriophages (Koonin, 1991).

Following the discovery and characterization of HCV, specific radioimmuno- and enzyme-linked assays for the detection of circulating HCV antibodies were developed by expression of a viral polypeptide in recombinant yeast (Kuo *et al.* 1989). c100-3 polypeptide, corresponding to nearly all of the

Figure 1.1

GENOME ORGANIZATION OF HCV AND HGV/GBV-C





E1, E2 = envelope proteins

NS2 = protease
NS3 = protease / RNA helicase
NS4 = cofactor
NS5B = RNA polymerase

NS4 protein of the HCV prototype isolate (Choo *et al.* 1989; Kuo *et al.* 1989; Choo *et al.* 1991) was used to coat the wells of microtitre plates, that were then incubated with human sera previously shown to transmit NANBH to chimpanzees (Kuo *et al.* 1989). Using this assay, antibodies to the c100-3 protein were detected in about 80% of sera from chronic, post-transfusion NANBH (PT-NANBH) patients from Italy and Japan. Antibodies were also present frequently in sporadic, community-acquired NANBH cases, such as patients from the USA with no identifiable source of parental exposure to the virus (Kuo *et al.* 1989). In addition, the immunoassay revealed a similar frequency of positive samples for linked donors and recipients, indicating that these antibodies were present in nearly all cases of PT-NANBH, and were particularly associated with chronic infections (Alter *et al.* 1989). As a result, it became apparent that HCV was the major cause of parenterally transmitted NANBH (Alter *et al.* 1989; Kuo *et al.* 1989).

1.2 NON-A,B,C,D HEPATITIS AGENTS

After the characterisation of HCV and the development of a specific screening assay for HCV antibodies, the use of molecular cloning techniques led to the successful isolation and identification of a cDNA clone of another human hepatitis virus designated hepatitis E virus (HEV) (Reyes *et al.* 1990). HEV is a non-enveloped single-stranded RNA virus which shares many biophysical (size, sedimentation, lack of a viral envelope) and biochemical features with

caliciviruses (Reyes *et al.* 1990). It was identified as a main cause of enterically-transmitted non-A, non-B hepatitis (Bradley, 1990; Reyes *et al.* 1990). Large "water-borne" epidemics of acute hepatitis in the Indian subcontinent, Central and South East Asia, parts of Africa, Central America were then found to be caused by HEV (Zuckerman, 1990; Velazquez *et al.* 1990).

More recently, virus particles of 27 to 37 nm diameter were detected by electron microscopy in liver samples from rhesus monkeys which had been inoculated with stool samples collected from patients with sporadic non-A, non-B hepatitis (Deka *et al.* 1994). The new virus, which was provisionally named hepatitis F virus (HFV), is a DNA virus with a genome of 20 kb (Deka *et al.* 1994). Research related to HFV is still at early stage and further investigations are necessary in order to examine its clinical importance in infectious hepatitis.

Despite these advances, the aetiology of about 20% cases of acute and community-acquired non-A, non-E hepatitis remained unknown, suggesting the existence of other causative agents (Alter *et al.* 1989; Aach *et al.* 1991). Further evidence for the existence of another potential infectious agent was provided by reports that 3 to 10% of patients with chronic hepatitis and cirrhosis were "cryptogenic" (Conn and Atterbury, 1993; Hammel *et al.* 1994; Kodali *et al.* 1994). The aetiology of these cases was unknown despite research into patient histories, histology, or the presence of serum markers of HBV or HCV infection, or evidence of autoimmune diseases. Results from extensive epidemiological studies provided additional evidence for this possibility. For example, in a five-year prospective study of acute NANB hepatitis in Greece in which 182 patients

were investigated, 47% of cases were established to be non-A,B,C,D related (Tassopoulos *et al.* 1992) with a small number of community-acquired NANBH cases associated with HEV infection (Tassopoulos *et al.* 1994). Earlier, biophysical and challenge studies in chimpanzees experimentally infected with factor VIII concentrate, suggested that there were two NANBH agents which induced separate and distinct episodes of NANBH (Bradley *et al.* 1983). Both agents could be recovered from liver tissue and plasma collected from a chimpanzee, one during the acute phase, and the other during the chronic phase of infection. One of the infectious agents proved to be chloroform sensitive and induced formation of cytoplasmic tubules, convoluted endoplasmic reticulum and dense reticular inclusion bodies in infected hepatocytes. In contrast, the second NANB agent was chloroform resistant and did not induce structural changes in hepatocytes.

1.3 DISCOVERY OF GB AGENTS

Early in 1967, the existence of a so-called "GB agent" was reported following experimental infection of laboratory animals (Deinhardt *et al.* 1967). Different serum samples from patients with unknown hepatitis were used to inoculate marmoset monkeys (*Saguinus sp.*). One of the inocula was obtained on the third day of jaundice from a 34-year old surgeon with acute hepatitis whose initials were GB. Five serial marmoset-to-marmoset passages were then carried out with this isolate. Elevation of liver enzyme levels and histological changes

in liver biopsies were observed in animals inoculated directly with GB serum and also in those inoculated during the serial passages. After the identification of HAV (Feinstone *et al.* 1973), further work was carried out to document the morphological characteristics of the GB agent (Almeida *et al.* 1976). Electron microscopy of the highly infectious serum collected from a marmoset infected with the GB agent identified aggregates of 20 to 22 nm particles, smaller than HAV particles which were about 27 nm in diameter (Almeida *et al.* 1976).

Passage and cross-challenge experiments of the GB agent in tamarins (*Saguinus labiatus*) were also carried out (Karayiannis *et al.* 1989). In these studies, intravenous inoculation of three tamarins (two of which were immune to HAV) with plasma containing the GB agent induced biochemical and histological changes consistent with acute hepatitis, indicating that GB and HAV are antigenically distinct. At the same time, the possibility that GB was an enterovirus shed in faeces was tested using extracts from faeces collected during the acute phase of the disease to inoculate two other tamarins. Neither of them developed hepatitis. This result indicated that GB is unlikely to be an enterovirus or related to enterally transmitted NANB hepatitis virus (Karayiannis *et al.* 1989). However, the detailed molecular characterization of the GB agent genome was only accomplished recently using a technique called representative differential analysis (RDA) (Lisitsyn and Wigler, 1993).

Nucleic acids were extracted from infected tamarin plasma (tester) and uninfected pre-inoculation plasma (driver). After cDNA synthesis by randomly-primed reverse transcription, second-strand synthesis was carried out using

random hexamers as primers. The double-stranded DNA products were digested with the restriction enzyme (*Sau*3A I). The digested tester DNA fragments were then ligated to the R *Bgl* primer set as initially described (Lisitsyn and Wigler, 1993) followed by hybridization of tester and driver DNAs. The selective amplification of tester-tester hybrids which were flanked by the R *Bgl* adaptor was performed by PCR using adaptor-specific primers. As a result, no tester-driver or driver-driver hybrids would be amplified and repeated rounds of hybridization and selective amplification are expected to lead to the selection of presumed exogenous nucleic acids.

Following the isolation of 76 clones, cross-hybridization experiments identified eleven unique clones, seven of which were examined extensively. Southern blot analysis failed to detect sequences from these clones in the genomes of human, uninfected tamarin, *Saccharomyces cerevisiae* or *Escherichia coli*. These sequences could only be detected after RT-PCR of samples from infectious tamarin plasma, suggesting that the corresponding clones were originally derived from an RNA molecule. Sequence analysis followed by searches of the GenBank and Swiss-Prot databases indicated that five of the seven clones had limited sequence similarity to the non-structural proteins of HCV. Further cloning revealed the existence in infected plasma of two unique plus-strand RNA genomes with similarity to flaviviruses, designated GBV-A and GBV-B, respectively (Simons *et al.* 1995b). More than 30 overlapping cDNA clones for each genome were isolated from serum and liver samples from a GB-infected tamarin using PCR-based techniques in order to generate the entire

nucleotide sequences of GBV-A and GBV-B (Muerhoff *et al.* 1995). GBV-A was found to contain 9,493 nucleotides while GBV-B comprises 9,143 nucleotides, each virus having a single ORF encoding potential polyproteins of 2,972 and 2,864 amino acids, respectively.

Comparison of the hydropathy profiles of GBV-A, GBV-B, HCV, bovine viral diahrroea virus (BVDV) and yellow fever virus (YFV) demonstrated that the polyproteins of these viruses had similar structures and organization although the overall amino acid sequence identity is low (Muerhoff *et al.* 1995). The polyproteins of GBV-A and GBV-B show limited identity to the helicase and RNA polymerase regions of HCV and other viruses encoding supergroup II helicases and polymerases (Koonin, 1991). Phylogenetic analysis of amino acid sequences within the helicase and RdRp domains revealed that GBV-A and GBV-B are neither genotypes of HCV nor genotypes of the same virus (Muerhoff *et al.* 1995) since the maximum distance between any two HCV genotypes for the RdRp region is 0.37 while the minimum distance calculated between GBV-A or GBV-B and any HCV genotype is 0.96. These results were supported by phylogenetic trees produced for either the helicase or RdRp sequence alignments of GBVs, HCV and other viruses in the *Flaviviridae* family; the GBVs were not present on the same major branch as the HCVs, indicating that GB agents could be classified into a new genera within the *Flaviviridae* (Muerhoff *et al.* 1995).

More transmission studies of the GB agents, together or separately, in tamarins were carried out in order to study the nature of GB virus infections

(Schlauder *et al.* 1995). Viraemia was monitored using the RT-PCR assay developed for detection of GBV-A and GBV-B RNAs, while antibody responses to GBV-A and GBV-B recombinant proteins were evaluated by enzyme-linked immunoassay (ELISA). The antigens were fragments of the putative NS3/NS4 and NS5 regions of GBV-A and fragments of the core, NS3/NS4 and NS5 regions of GBV-B. Antibodies to GBV-B epitopes were detected in serum samples collected from tamarins inoculated with both GB viruses or GBV-B alone but there were no detectable antibody responses to the GBV-A epitopes in any of the animals. GBV-B viraemia, detected by RT-PCR, caused an elevation of ALT after inoculation of GBV-B, either alone or together with GBV-A. On the contrary, no enzyme elevation was observed in tamarins with only GBV-A viraemia. These results indicated that GBV-B causes hepatitis in inoculated tamarins (Schlauder *et al.* 1995). However, the peak levels of ALT in GBV-B infected animals was lower than in GBV-A/GBV-B coinfected tamarins suggesting that the severity of hepatitis is related to the presence of both GB agents.

In order to examine the seroprevalence of antibodies to GBV-A and GBV-B in humans, serum samples collected from different individuals were screened for GB viruses (Simons *et al.* 1995a) using ELISA assays for recombinant proteins of GBV-A and GBV-B (Schlauder *et al.* 1995). Three distinct groups were examined: volunteer blood donors previously tested negative for HBsAg, anti-HCV and anti-HBc; intravenous drug users (IVDUs) with a high prevalence of antibody to HCV (99%) and HBV (76%); and a group from

Western Africa where infection with agents of viral hepatitis is relatively common. Amongst volunteer blood donors, a seroprevalence of 0.3% (3/860) was observed for antibodies to GBV-A recombinant proteins and 1.2% (12/960) for GBV-B recombinant proteins. In contrast, screening of the IVDU group showed the presence of antibodies to GBV-A and GBV-B in 14%. A similar seroprevalence was obtained in the samples from West Africa where antibodies to GBV-A were detected in 8.4% and to GBV-B in 14.6% with 3.07% of samples positive for both viruses (Simons *et al.* 1995a).

These immunoreactive sera were then tested for the presence of viral sequences by RT-PCR using degenerate primers based on GBV-A, GBV-B and HCV sequences from the NS3 (helicase) region. An amplification product was obtained from a West African sample, and this had limited nucleotide sequence identity to GBV-A (59%), GBV-B (47.9%) and HCV-type 1a (57.3%). Further sequence analysis and alignment of the predicted translation product with sequences of other members of the *Flaviviridae* family suggested that this sequence belonged to a new virus. Because of the relatively high degree of identity with GBV-A at both nucleotide and amino acid levels (59% and 64%, respectively), this virus was named GB virus C (GBV-C) (Simons *et al.* 1995a).

No PCR product was obtained when the RT step was omitted, and GBV-C sequences could not be detected in DNA samples from human, Rhesus monkey, *Saccharomyces cerevisiae* or *Escherichia coli*. These results confirmed that the GBV-C sequences were derived from an RNA virus. Finally, sequence analysis of regions up- and downstream of the helicase region obtained from

overlapping genomic clones revealed that the GBV-C genome is 9,125 nucleotides in length with an ORF of 2,906 amino acids (Leary *et al.* 1996b).

Serological studies of the prevalence of GBV-C were initiated by screening serum samples obtained from individuals diagnosed with non-A-E hepatitis (Simons *et al.* 1995a). For this purpose, fragments of the virus genome were expressed as recombinant proteins in *E. coli* and then used to detect immunoreactive GBV-C sera by ELISA. From a total of 161 samples, 5 (3.1%) were immunoreactive to GBV-C proteins while 26 (16.1%) were positive for recombinant proteins of GBV-A, GBV-B or GBV-C. Eight out of these 26 sera were positive by RT-PCR using GBV-C specific primers. The samples originated from individuals with distinct clinical background: one patient was co-infected with HCV, another patient had hepatitis-associated aplastic anaemia, two patients had acute hepatitis without markers for infection with hepatitis A-E viruses, and the four remaining patients were Africans for whom no clinical information was available. The presence of GBV-C RNA in individuals with unknown hepatitis suggested, for the first time, that the GBV-C might be one of the causative agents of nonA-E hepatitis. However, since GBV-C and HCV co-infected one patient, there may have been common risk factors for acquiring these infections, and it was not clear whether the hepatitis symptoms were due to HCV, GBV-C or both (Simons *et al.* 1995a).

In parallel, a search for potential new infectious agents associated with nonA-E hepatitis cases by another group of scientists led to the identification and characterisation of a novel RNA virus designated hepatitis G virus (HGV) (Linnen *et al.* 1996). The source of this virus was plasma from an American patient (PNF2161) with chronic NANB hepatitis. The patient was initially found to be negative for HCV using a first-generation immunoassay, but was positive by second-generation HCV immunoassay and also for HCV RNA by RT-PCR using HCV 5'UTR-specific primers (Linnen *et al.* 1996).

RNA extracted from the plasma sample and reverse-transcribed with random primers was amplified by sequence-independent single primer amplification (SISPA) followed by cloning of the amplified products into the *E. coli* expression vector λ gt11. The same plasma was then used to screen colonies from the cDNA library, and a single immuno-reactive clone was identified and sequenced. Using primers derived from the clone sequence, PCR amplification was performed on various templates such as human genomic DNA, *E. coli* DNA and *S. cerevisiae* DNA. The absence of any PCR products in these reactions indicated that the immuno-reactive clone contained exogenous sequences. No PCR product was obtained from the plasma sample when reactions were carried out without a RT step, demonstrating that the sequence derived from RNA. Anchored PCR was used to generate multiple overlapping clones whose combined sequence 9,392 nucleotides encoding a putative viral polyprotein of

2,873 amino acids. This new RNA virus was designated hepatitis G virus (HGV) (Linnen *et al.* 1996).

The genome organization of HGV resembles that of HCV (e.g. structural genes located towards the 5'end and non-structural genes towards the 3'end of the genome), indicating the existence of a flavivirus-like RNA genome. Comparison of the predicted amino acid sequence of the HGV polyprotein to other flaviviruses showed limited identity to HCV type 1 (26.8%), GBV-B (28.4%) or GBV-A (43.8%), indicating that HGV is not a new HCV genotype (Linnen *et al.* 1996). A second HGV sequence of 9,103 nucleotides was isolated from the plasma of an asymptomatic individual (R10291) with a history of intermittent elevations in liver enzyme concentrations (Linnen *et al.* 1996). The sample was serologically negative for both HBV and HCV and was also HCV negative by RT-PCR. The sequence identity between PNF2161 and R10291 isolates was 90.5% at nucleotide level and 97.5% at amino acid level. However, comparison of HGV and GBV-C indicated that the two isolates were 85.5% nucleotide identical and 100% amino acid identical in the NS3 helicase region (Linnen *et al.* 1996). The suggestion that HGV and GBV-C are independent isolates of the same virus was subsequently confirmed by comparison of their entire genomes which revealed amino acid and nucleotide sequence identities of 95% and 85%, respectively (Zuckerman, 1996). The terminology HGV/GBV-C will be used throughout this study.

1.5 GENOME ORGANIZATION OF HGV/GBV-C

5'-UNTRANSLATED REGION

HGV/GBV-C genome is a single-stranded positive sense RNA molecule of approximately 9,400 nucleotides with a single long open reading frame (ORF) flanked by untranslated regions at both its 5' and 3'ends (Figure 1.1) (Simons *et al.* 1995a; Leary *et al.* 1996a). The 5'-untranslated region (5'-UTR) is relatively long (553 nucleotides) and has limited sequence identity to the 5'-UTR of GBV-A but not with the 5'-UTRs of HCV, GBV-B or pestiviruses (Simons *et al.* 1996). Because this region contains many nucleotide deletions/insertions and substitutions, no consistent ORF exists upstream of the E1 gene that would be capable of encoding a core-like protein (Erker *et al.* 1996; Muerhoff *et al.* 1996; Okamoto *et al.* 1997; Takahashi *et al.* 1997; Wang *et al.* 1997).

Although up to five potential in-frame AUG codons are present within the 5'-UTR, the AUG codon located immediately upstream of the putative signal sequence for E1 protein is considered to be the site of translation initiation (Simons *et al.* 1996). This conclusion was reached following *in vitro* transcription translation (IVTT) reactions of mono-cistronic plasmids containing 5'-terminal fragments of the HGV/GBV-C with artificially introduced mutations and sequencing of the protein products by Edman degradation. In addition, this AUG codon which is situated downstream of an oligopyrimidine tract is conserved amongst all HGV/GBV-C isolates and is in - frame with the putative

HGV/GBV-C polyprotein (Muerhoff *et al.* 1996b; Simons *et al.* 1996b; Pickering *et al.* 1997b; Takahashi *et al.* 1997b), suggesting that the translation initiation in HGV/GBV-C RNA is achieved through an internal ribosome entry mechanism. Internal entry of ribosomes is observed for HCV (Tsukiyama Kohara *et al.* 1992), pestiviruses (Poole *et al.* 1995) and picornaviruses (Jackson *et al.* 1994), and is dependent on the existence of an internal ribosome entry site (IRES) or ribosome landing pad (RLP) represented by a defined fragment of the 5'-UTR sequence to which the 40S ribosome unit can bind in order to initiate 5'cap-independent protein synthesis. The presence of IRES elements within HGV/GBV-C 5'-UTR was demonstrated using bicistronic vectors containing the 5'-UTR fragment cloned between bacterial chloramphenicol acetyltransferase (CAT) gene and luciferase (Luc) gene in IVTT reactions followed by analysis of the protein products (Simons *et al.* 1996). However, under the experimental conditions tested, the HGV/GBV-C IRES seems to have a very low activity (2-5%) compared with the HCV IRES.

Thermodynamic and phylogenetic analysis of HGV/GBV-C 5'-UTR sequences from different isolates predict that this region folds into an RNA secondary structure (Simons *et al.* 1996b; Pickering *et al.* 1997b), with structural elements that are similar to the 5'-UTR of GBV-A but distinct from those predicted for HCV (Simons *et al.* 1996).

3'-UNTRANSLATED REGION

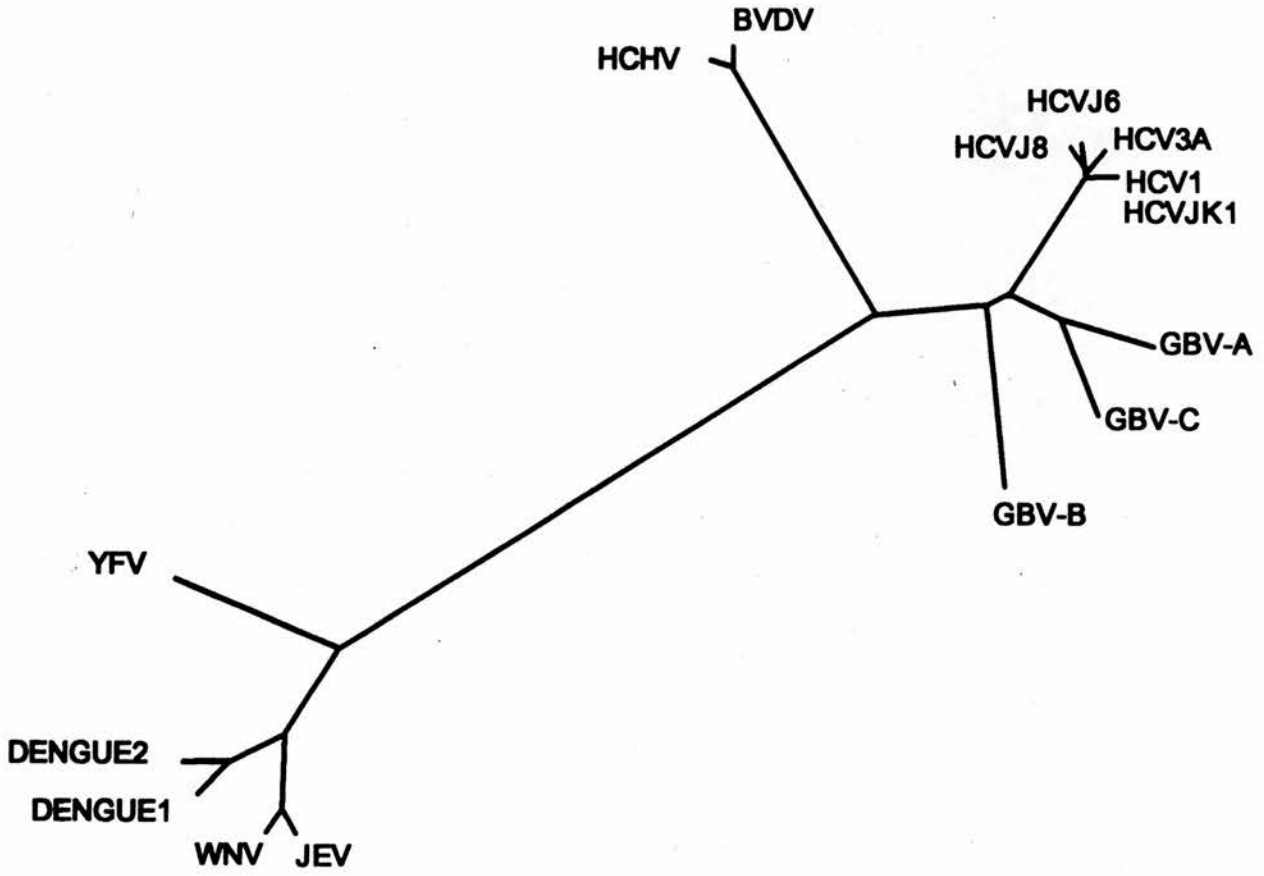
The 3'-untranslated region (3'-UTR) of the HGV/GBV-C genome is characterised by a high degree of sequence conservation with a similarity between isolates of 94.6% to 98.4% (Erker *et al.* 1996; Okamoto *et al.* 1997), and there is evidence for a conserved RNA secondary structure (Erker *et al.* 1996; Okamoto *et al.* 1997; Katayama *et al.* 1998) . Unlike HCV, no poly(U) or poly(A) tail has been identified at the 3'terminus of the HGV/GBV-C genome (Erker *et al.* 1996; Leary *et al.* 1996; Linnen *et al.* 1996; Nakao *et al.* 1997; Okamoto *et al.* 1997; Katayama *et al.* 1998).

CODING REGIONS

Like HCV and other members of the *Flaviviridae*, the ORF of HGV/GBV-C encodes a single polyprotein with the structural genes located towards the amino-terminus and non-structural genes containing helicase, protease and RNA-dependent RNA polymerase (RdRp) motifs towards the carboxyl-terminus. Phylogenetic analysis of NS3 helicase region from various members of *Flaviviridae* family demonstrated that HGV/GBV-C is more closely related to GBV-A than GBV-B or any of HCV genotypes (Figure 1.2) (Simons *et al.* 1995a).

The precursor polyprotein of HGV/GBV-C is post-translationally cleaved by a combination of host and viral proteases into structural and non-structural proteins (Leary *et al.* 1996c). Comparisons with HCV have identified

Figure 1.2 Phylogenetic tree based on analysis of the putative RNA helicase region from members of *Flaviviridae* family-Phylip method (Simons *et al*, 1995)



homologues of the putative E1 and E2 envelope proteins as well as conserved eukaryotic signal sequence-like motifs at cleavage sites between the E1/E2 and E2/NS2 genes (Erker *et al.* 1996; Leary *et al.* 1996; Takahashi *et al.* 1997). An Asn-Cys-Cys motif is located near the amino-terminus of the putative E1 protein in GBV-A, GBV-B, HGV/GBV-C and HCV-1 (Erker *et al.* 1996; Leary *et al.* 1996; Takahashi *et al.* 1997). Other conserved sites are a potential N-linked glycosylation site in the E1 region and three in the E2 region, suggesting that HGV/GBV-C envelope proteins are glycosylated during cellular processing (Leary *et al.* 1996; Okamoto *et al.* 1997; Takahashi *et al.* 1997; Wang *et al.* 1997). However, there are fewer glycosylation sites than for HCV in which E1 and E2 have five-six or eight-eleven N-linked glycosylation sites, respectively (Miyamura and Matsuura, 1993). Another difference between HCV and HGV/GBV-C concerns the E2 gene which for HCV represents the most variable region of the genome. Domains of extreme hypervariability (HVR) have been identified (Hijikata *et al.* 1991a; Weiner *et al.* 1991a) with HVR-1, representing the amino-terminal 25-30 amino acids of the E2 protein, being the most variable and intensively studied region. The HVR of HCV-1 is thought to play a role in HCV neutralisation (Taniguchi *et al.* 1993; Kojima *et al.* 1994). In contrast, sequence analysis of HGV/GBV-C structural proteins reveals lower degree of variability within the envelope proteins (at most 10%), with no evidence of "hypervariable" regions, although amino acid substitutions are often present at the amino-terminus of E2 (Erker *et al.* 1996; Takahashi *et al.* 1997; Wang *et al.* 1997). It was even suggested that the amino-terminus of HGV/GBV-C E2 which contains

two consecutive amino acid positions that are highly variable, could be exposed on the virion surface and be recognised by antibodies (Takahashi *et al.* 1997). However, other studies have failed to identify hypervariable regions in the E2 gene of HGV/GBV-C (Lim *et al.* 1997; Kato *et al.* 1998).

Another difference between the genome organization of HGV/GBV-C and HCV is that the HGV/GBV-C genome does not appear to encode a nucleocapsid or core protein. In the case of HCV, the viral core protein is located at the amino-terminus of the polyprotein (Choo *et al.* 1991), being released from it by proteolytic cleavage mediated by host signal peptidase (Hijikata *et al.* 1991b). However, analysis of sequences upstream of the putative E1 gene obtained from various HGV/GBV-C isolates revealed amino acid fragments of variable length and composition, and sometimes absent (Erker *et al.* 1996; Linnen *et al.* 1996; Muerhoff *et al.* 1996; Okamoto *et al.* 1997; Takahashi *et al.* 1997; Wang *et al.* 1997; Katayama *et al.* 1998). This observation suggests that HGV/GBV-C may assemble in a capsid-free particle or by using a host-encoded protein or a protein encoded by a different virus (Takahashi *et al.* 1997; Theodore and Lemon, 1997).

The degree of similarity between the non-structural domains of HGV/GBV-C and HCV is much higher. For example, the Zinc-protease motif, identified originally in the NS2 gene of HCV (Grakoui *et al.* 1993), was also found in HGV/GBV-C (Linnen *et al.* 1996). The HCV NS2 protein was shown to play a role in NS2/NS3 cleavage with its catalytic activity being affected by mutations of certain amino acids (His_{952} and Cys_{993}) (Grakoui *et al.* 1993). The

amino acids essential for the protease activity of HGV/GBV-C NS2 have been identified by mutation analysis using a baculovirus expression system (Belyaev *et al.* 1998). Following infection of Sf21 insect (*Spodoptera frugiperda*) cells with constructs either containing wild type (wt) NS2 protease or carrying mutations His_{849}→Tyr or Cys_{890}→Leu in the NS2 active site. Cell proteins were separated by SDS-PAGE and analysed by Western blotting with rabbit antisera raised against the HGV/GBV-C NS2 peptide. In the wt construct, NS2 was efficiently cleaved from NS3 since NS2 was the only product reacting with anti-NS2 antibodies, and there was no protein-associated reactivity at higher molecular mass (Belyaev *et al.* 1998). In contrast, no free NS2 and NS3 products were obtained on immunoblots of protein expressed from mutated constructs, indicating that NS2 protein activity was required for the NS2/NS3 cleavage.

As in HCV, the carboxy-terminus of the HGV/GBV-C NS3 region encodes a polypolypeptide containing amino acid residues that are conserved in the helicases of supergroup II positive-strand RNA viruses (Leary *et al.* 1996b). In particular, the amino-terminus contains a catalytic triad composed of Histidine, Aspartic acid and Serine which are specific to chymotrypsin-like serine proteases (Leary *et al.* 1996c; Linnen *et al.* 1996c). The proteolytic activity of HCV NS3 protein was shown to be required for cleavages between NS3/NS4A, NS4A/NS4B, NS4B/NS5A and NS5A/NS5B in order to release the corresponding mature proteins (Bartenschlager *et al.* 1993; Grakoui *et al.* 1993; Lin *et al.* 1994; Manabe *et al.* 1994). Similarly, site-directed mutagenesis of the catalytic site serine of HGV/GBV-C NS3 has demonstrated that this serine

protease is essential for cleavage of NS3/NS4, NS4/NS5A and NS5A/NS5B (Belyaev *et al.* 1998). Efficient cleavage at NS4B/NS5A was dependent on the presence of the NS4A protein as a cofactor for NS3 activity, as previously described for HCV (Belyaev *et al.* 1998). Identification of the NS4A active domain was achieved by making various truncations at both ends of the NS4A region in recombinant baculovirus constructs and analysing the ability of these constructs to complement the NS3-mediated cleavage of NS4B/NS5A. Western blot analysis of NS4B/NS5A cleavage products using antibodies to the NS4B peptide showed that the domain critical for NS4A activity is located between Leu_{1561} and Ala_{1598} (numbered as in PNF2161 sequence) (Belyaev *et al.* 1998).

The similarity between HCV and HGV/GBV-C polyproteins also extends throughout the NS5B region where a number of amino acids, which are conserved in other *Flaviviridae* members, have been identified in the HGV/GBV-C NS5B domain (Leary *et al.* 1996c; Linnen *et al.* 1996c). In particular, the amino acid motif Gly-Asp-Asp located within the NS5B region is characteristic of all known RNA-dependent RNA polymerases (Kamer and Argos, 1984) and so, HGV/GBV-C NS5B is presumed to be the viral polymerase (Leary *et al.* 1996c; Linnen *et al.* 1996c).

1.6 DETECTION OF HGV/GBV-C

At present, detection of HGV/GBV-C is based on nucleic acid amplification by RT-PCR. Degenerate oligonucleotide PCR primers designed from a consensus sequence of NS3 helicase genes of HCV, GBV-A and GBV-B were initially used to isolate GBV-C RNA from serum samples of eight patients, of whom four had cryptogenic hepatitis (Simons *et al.* 1995a). Comparison of the nucleotide sequences from these isolates indicated that they were 83-86% identical. In order to improve the detection of HGV/GBV-C, degenerate PCR primers for the NS3 region were designed based on a consensus sequence of these eight isolates (Leary *et al.* 1996a). The primers were more efficient than the original set when used in a "touchdown" PCR assay to detect HGV/GBV-C sequences in human clinical samples from three different groups: non-A, non-E hepatitis cases, individuals with indeterminant anti-HCV and IVDUs. For example, in case of 62 HCV indeterminants (samples seropositive for a single HCV marker (NS5) but negative for other antigens), the RT-PCR for HGV/GBV-C RNA indicated that 6.5% of samples were positive with the original set (Simons *et al.* 1995a), while 16.1% were positive with the consensus primer pair (Leary *et al.* 1996a). However, nucleotide variation in coding regions means that the use of primers derived either from NS3 (Simons *et al.* 1995a; Leary *et al.* 1996a) or NS5A regions (Linnen *et al.* 1996) of HGV/GBV-C genome could reduce the sensitivity of PCR assay since there might be mismatches between primers and the cDNA template (Simons *et al.* 1995a).

The RT-PCR-based detection of HGV/GBV-C has therefore been optimised by identifying domains with a high degree of sequence conservation. Nucleotide sequence alignments of 5'-UTR fragments obtained from different individuals revealed the presence of such localized conserved domains (Jarvis *et al.* 1996; Muerhoff *et al.* 1996; Schlueter *et al.* 1996; Chayama *et al.* 1997) and comparison of 5'-UTR primer sets with NS3 (Jarvis *et al.* 1996; Muerhoff *et al.* 1996; Chayama *et al.* 1997) or NS5A primer sets (Schlueter *et al.* 1996) suggested that they increase the sensitivity of HGV/GBV-C RNA detection in nested PCR assays. For example, of 12 samples found to be HGV/GBV-C positive using several different 5'-UTR primer sets, only 11 were positive using NS3 primers (Muerhoff *et al.* 1996a). Similarly, HGV/GBV-C RNA was detected in 8 out of 50 Japanese patients with hepatocellular carcinoma (16%) using 5'-UTR primers but only in 5 out of 50 (10%) when samples were tested with NS3-derived primers (Chayama *et al.* 1997).

Efforts have also been made to develop specific immunoassays for the identification of individuals showing an immune response to HGV/GBV-C infection (PilotMatias *et al.* 1996; Dille *et al.* 1997; Tacke *et al.* 1997). In contrast to PCR-based assays, the antibody prevalence to HGV/GBV-C has proved difficult to assess. Initially, attempts were made to identify immuno-reactive epitopes within HGV/GBV-C viral polyprotein (PilotMatias *et al.* 1996). Regions covering the large ORF of HGV/GBV-C genome were cloned and then expressed in *E.coli* as fusion proteins. These proteins were examined for immuno-reactivity by Western blot analysis using sera from GB agent-infected

tamarins, original GB serum and sera from different individuals considered at risk for parenteral exposure. Although none of the HGV/GBV-C fragments were reactive with either GB-infected tamarin serum or GB serum, fragments spanning NS3-NS5B region were reactive with one or the other of two sera from individuals with or at risk for nonA-E hepatitis (PilotMatias *et al.* 1996). Epitopes within these potential immuno-reactive fragments were mapped by generating small overlapping clones by PCR and expressing them as fusion proteins to be tested by Western blot. As a result, four specific regions within the NS3, NS4A/B and NS5A/B proteins of HGV/GBV-C were found to be immuno-reactive although reactivity with some of the sera tested was weak and inconsistent (PilotMatias *et al.* 1996).

Similar results were obtained in another study (Dawson *et al.* 1996) in which three HGV/GBV-C recombinant proteins spanning NS3 through a portion of NS5B region (PilotMatias *et al.* 1996) and two additional ones encoding E2 and NS5A gene fragments respectively, were used as antigens in ELISA assays to detect the presence of antibodies in samples which were PCR-positive for HGV/GBV-C RNA. Of 43 samples, 11 (25.6%) were antibody-positive for HGV/GBV-C proteins, but no antigen was consistently recognised and, in almost all cases, each serum was reactive with only one of the recombinant proteins (Dawson *et al.* 1996). However, the prokaryotically expressed HGV/GBV-C recombinant proteins used as antigens in these ELISAs did not represent their native conformation since they were not glycosylated (in case of E2 fragment) or they had abnormal amino or carboxyl termini. These modifications could

significantly affect the secondary structure of the proteins and, so, alter conformational epitopes, possibly explaining the lack of correlation between HGV/GBV-C RNA detection and an antibody response to HGV/GBV-C proteins.

Two research groups investigated the possibility of developing an ELISA assay specific for antibodies to HGV/GBV-C E2 protein (Dille *et al.* 1997; Tacke *et al.* 1997) following a similar approach used to validate the presence of antibody to the HCV E2 protein as a useful marker of HCV infection (Lesniewski *et al.* 1995). Recombinant plasmids containing a DNA fragment encoding the HGV/GBV-C E2 protein were transfected into dihydrofolate reductase deficient Chinese hamster ovary (CHO) cells (Dille *et al.* 1997a; Surowy *et al.* 1997a; Tacke *et al.* 1997a). Two approaches have been carried out in designing the expression construct. In the first report, the E2 fragment was cloned downstream from a sequence encoding an eight-amino acid peptide (FLAG) (Tacke *et al.* 1997a). After expression, solubilised protein was bound to a streptavidin-coated ELISA plates using the biotin-conjugated FLAG-specific antibody M1. The plates were then incubated with samples of human sera and antibodies to the HGV/GBV-C E2 protein detected using anti-human IgG-peroxidase conjugate (Tacke *et al.* 1997a). In the second approach, the HGV/GBV-C E2 sequence was cloned into a plasmid so as to encode a carboxy-terminal truncated form of the protein in which a hydrophobic tail that might prevent secretion was replaced with the FLAG sequence (Dille *et al.* 1997; Surowy *et al.* 1997). After expression, the E2 protein was purified on an

immuno-affinity column containing anti-FLAG monoclonal antibody. Purified E2 protein was then coated onto polystyrene beads and dilutions of human sera were tested for the presence of specific antibodies (Dille *et al.* 1997; Surowy *et al.* 1997). Treatment of purified protein with N-Glycanase under both denaturing and native conditions indicated that the secreted protein was N-glycosylated while amino-terminal amino acid sequencing confirmed that the proper protein was expressed (Surowy *et al.* 1997).

These assays were used to screen serum samples from population groups such as volunteer blood donors, IVDUs and nonA-E hepatitis patients (Dille *et al.* 1997; Surowy *et al.* 1997; Tacke *et al.* 1997). Surprisingly, antibodies to HGV/GBV-C E2 protein were detected in individuals with a high risk for parenterally-transmitted viruses but who were not viraemic (PCR-negative for HGV/GBV-C RNA). For example, the prevalence of antibodies to E2 was relatively low (3.0%-8.1%) in volunteer blood donors but was higher in IDVUs (85.2%), and West African residents (13.3%), but all were PCR-negative for HGV/GBV-C RNA (Dille *et al.* 1997). Similar findings were reported in other studies and suggest that a humoral immune response to the HGV/GBV-C E2 protein is associated with viral clearance (Surowy *et al.* 1997; Tacke *et al.* 1997). As a consequence and unlike HCV infection where anti-E2 antibodies and RNA occur together (Lesniewski *et al.* 1995), antibodies to HGV/GBV-C E2 can be considered as a marker for recovery from HGV/GBV-C infection.

The specificity of these E2 immuno-assays was assessed by testing sequential samples obtained from eleven patients with post-transfusion hepatitis,

seven of which were PCR-positive for both HCV RNA and HGV/GBV-C RNA (Tacke *et al.* 1997a). All the patients were initially anti-E2 negative, while four individuals developed antibodies to HGV/GBV-C E2 during the following year, two of whom became HGV/GBV-C RNA negative while no cross-reacting antibodies to HCV E2 could be detected.

An improved ELISA for the detection of antibodies to HGV/GBV-C E2 is based on the use of monoclonal antibodies to HGV/GBV-C E2 as capture antibodies (Tacke *et al.* 1997b) instead of anti-FLAG monoclonal antibodies (Tacke *et al.* 1997a). With this assay, several sera in a panel of 16 post-transfusion patients followed up for 16 years were found to be HGV/GBV-C E2 positive that had been found to be anti-E2 negative with previous assay. However, an inverse correlation was still observed between the presence of viral RNA and antibodies to E2. In addition, this study suggests that immunity to HGV/GBV-C may be long-lasting since circulating antibodies to E2 could still be detected 14 years after seroconversion (Tacke *et al.* 1997b).

1.7 EPIDEMIOLOGY

1.7.1 PREVALENCE OF HGV/GBV-C INFECTION

Since the discovery and characterisation of HGV/GBV-C virus, there have been numerous studies of the prevalence of HGV/GBV-C in different population groups based on RT-PCR assays using primers derived from different regions of the viral genome.

The prevalence of viraemia amongst healthy blood donors from various countries of the world ranges from 0.5% to 1.7% (Dawson *et al.* 1996; Linnen *et al.* 1996; Orito *et al.* 1996; Feucht *et al.* 1997; Wang *et al.* 1997). Higher rates have been found in British (3.2%) (Jarvis *et al.* 1996), Australian (4%) (Moaven *et al.* 1996a) and Vietnamese blood donors (7.4%) (Brown *et al.* 1997a). In contrast, the rate of active infection with HCV in similar populations was much lower, at between 0.01-1% (Jarvis *et al.* 1996; Linnen *et al.* 1996; Wang *et al.* 1997). These data imply that HGV/GBV-C would be present in blood products prepared from large plasma pools such as clotting factor concentrates and immunoglobulin batches (Jarvis *et al.* 1996; Linnen *et al.* 1996; Nubling and Lower, 1996; Feucht *et al.* 1997; Sheng *et al.* 1997). Investigations revealed that HGV/GBV-C RNA was present in 94% of non-virus-inactivated factor VIII concentrates compared to 6% of heat- or solvent-detergent-treated concentrates (Jarvis *et al.* 1996) and in 7-40% of commercial plasma pools used for production of blood products (Nubling and Lower, 1996). Similarly, 33-44% of immunoglobulin preparations were found to be contaminated with HGV/GBV-C (Jarvis *et al.* 1996; Nubling and Lower, 1996), although only 4 of 57 (7%) immunoglobulin recipients were found to be positive for viral RNA (Jarvis *et al.* 1996). The presence of HGV/GBV-C RNA has also been investigated in serum samples from haemophiliacs since these patients have been exposed to unheated and/or dry heated pooled clotting factor concentrates, and have been found to have high frequencies of HCV infection, and to be at risk of developing chronic hepatitis (Watson *et al.* 1992). Studies have revealed that 14-35% of

haemophiliacs treated with non-virus-inactivated concentrates were positive for HGV/GBV-C RNA (Jarvis *et al.* 1996; Linnen *et al.* 1996; Feucht *et al.* 1997; Sheng *et al.* 1997) while 83% were found to be infected with HCV (Jarvis *et al.* 1996). The difference in the prevalence of HCV and HGV/GBV-C in haemophiliacs suggests that recovery from HGV/GBV-C infection could be a more frequent event than recovery from HCV infection (Jarvis *et al.* 1996). This hypothesis is supported by later studies reporting that the presence of antibodies to the HGV/GBV-C E2 protein, considered as a marker of recovery from viraemia (Tacke *et al.* 1997a), are detected in 58% of haemophiliacs (Karayiannis *et al.* 1997b).

A high rate of HGV/GBV-C viraemia was also observed amongst other groups of individuals with risk factors for viral transmission by blood such as IVDUs (16-49%, with 54-100% coinfected with HCV) (Dawson *et al.* 1996; Linnen *et al.* 1996; Schreier *et al.* 1996; Diamantis *et al.* 1997; Feucht *et al.* 1997), haemodialysis patients (3.1-55%, with 44-81% coinfected with HCV) (Masuko *et al.* 1996; Tsuda *et al.* 1996; Feucht *et al.* 1997; Wang *et al.* 1997) and multiply transfused patients with aplastic anaemia (Simons *et al.* 1995b; Brown *et al.* 1997b). The virus can produce persistent infections since HGV/GBV-C RNA remained detectable in haemodialysis patients for up to 16 years (Masuko *et al.* 1996).

However, these epidemiological studies based on RT-PCR assays are limited because they only give information about individuals or groups with HGV/GBV-C viraemia. Serological tests indicate that the appearance of an

antibody response to the envelope protein E2 of HGV/GBV-C generally coincide with the decline and loss of viraemia (Dille *et al.* 1997a; Tacke *et al.* 1997a). In addition, population surveys revealed that the prevalence of anti-E2 was higher than that of HGV/GBV-C RNA in various groups (Dille *et al.* 1997; Gutierrez *et al.* 1997; Lou *et al.* 1997; Ross *et al.* 1998). As a result, in order to evaluate the total exposure to HGV/GBV-C infection in a population, assays for both HGV/GBV-C RNA and HGV/GBV-C E2 antibody must be carried out. In European countries, anti-E2 seropositivity ranges from 10.9% (Germany) to 15.3% (Austria) while higher rates have been recorded in South Africa (20.3%) and Brazil (19.5%) (Ross *et al.* 1998). Total exposure, considered as a sum of HGV/GBV-C RNA positive and anti-E2 positive samples, has been estimated at between 5.5% (Gutierrez *et al.* 1997) and 8-13% (Lou *et al.* 1997) in American volunteer blood donors and even higher in commercial blood donors (40.5-50%) (Gutierrez *et al.* 1997; Lou *et al.* 1997), IVDUs (89.2%) and acute HCV patients (70.7%) (Gutierrez *et al.* 1997). At the same time, the fact that the exposure rate is higher than the active infection rate (e.g. 3 to 6 times higher in volunteer donors) suggests that more than half of those infected with HGV/GBV-C eventually clear the virus (Dille *et al.* 1997; Gutierrez *et al.* 1997; Tacke *et al.* 1997). However, more studies are necessary to assess whether antibodies to HGV/GBV-C E2 are protective against HGV/GBV-C reinfection.

1.7.2 TRANSMISSION OF HGV/GBV-C

A conclusion common to these different epidemiological studies is that there is an association between the presence of HGV/GBV-C RNA and exposure to blood or blood products. A parenteral route of transmission was first suggested by the observation that transfusion recipients negative for HGV/GBV-C RNA prior to transfusion became HGV/GBV-C RNA positive afterwards (Linnen *et al.* 1996). In a larger study of 400 adults who underwent cardiac surgery, transfusion-related HGV/GBV-C infection was identified in 9% of patients (Wang *et al.* 1996). In addition, a high rate of HCV and/or HBV coinfection has been identified in HGV/GBV-C positive individuals emphasizing the common parenteral mode of transmission for these viruses. For example, the presence of HGV/GBV-C RNA was reported in 5.6-24.4% of chronic HCV patients (Dawson *et al.* 1996; Linnen *et al.* 1996; Tanaka *et al.* 1996; Feucht *et al.* 1997; Kao *et al.* 1997; Saiz *et al.* 1997; Pawlotsky *et al.* 1998), in 3.2-9% of chronic HBV carriers (Linnen *et al.* 1996; Kao *et al.* 1997) and in 18.2-32% of individuals with HIV infection (Feucht *et al.* 1997; Fiordalisi *et al.* 1997).

However, the higher prevalence of HGV/GBV-C RNA in volunteer blood donors than that of HCV suggested that alternative routes of viral transmission may be responsible for the maintenance of virus at this rate in the general population (Moaven *et al.* 1996; Chen *et al.* 1997; Fischler *et al.* 1997; Rubio *et al.* 1997; Scallan *et al.* 1998). Mother-to-baby transmission was examined as a potential mechanism of HGV/GBV-C infection. Preliminary evidence for this

hypothesis was provided by a case of a mother who was HGV/GBV-C RNA positive before and at delivery and whose baby tested HGV/GBV-C RNA negative at birth but was positive at four and six weeks of age (Moaven *et al.* 1996b). Other studies involving follow-up of children of pregnant women, considered as being at risk for parenterally-transmitted viruses, showed that HGV/GBV-C infection was transmitted efficiently from mother to baby and that the rate of transmission was much higher than that of HCV or HIV-1 (Feucht *et al.* 1996; Zanetti *et al.* 1998). A similar study of women coinfected with HCV and HGV/GBV-C suggested that HGV/GBV-C transmission from mother to baby can occur independently of HCV transmission (Fischler *et al.* 1997). None of these studies described any clinical or biochemical signs of liver disease in HGV/GBV-C infected babies (Feucht *et al.* 1996; Fischler *et al.* 1997; Zanetti *et al.* 1998).

The possibility of sexual transmission of HGV/GBV-C has also been investigated. HGV/GBV-C RNA was detected in 13.9-21% of prostitutes (Rubio *et al.* 1997; Wu *et al.* 1997) and in 13.4% of homosexual men (Rubio *et al.* 1997). The prevalence of HGV/GBV-C infection was significantly higher in heterosexual partners of HGV/GBV-C infected index cases (individuals with elevated ALT levels and/or at risk for parenterally/sexually transmitted viral disease) than in HGV/GBV-C negative index patients (21.7% versus 6.2%) (Rubio *et al.* 1997). More evidence for the spread of virus by the sexual route was revealed in a recent study in which the combined frequencies of HGV/GBV-C viraemia and anti-E2 antibody were measured and used to estimate total

exposure to the virus of individuals at risk for sexually transmitted diseases (Scallan *et al.* 1998). The frequency of total exposure was higher in prostitutes (40%) compared to the control group (5%) while amongst homosexual men, the total exposure to the HGV/GBV-C in HIV-positive individuals was 63% compared with 37% in HIV-negative individuals. A higher frequency of persistent infection was also observed among HIV-positive male homosexuals from whom retrospective samples were available since 7 of 11 retrospective samples were PCR positive for HGV/GBV-C.

HGV/GBV-C RNA has been detected in saliva samples of some infected individuals (Chen *et al.* 1997; Blair *et al.* 1998). Two saliva samples and six serum samples collected from 34 patients with chronic HCV infection were found to be positive for HGV/GBV-C RNA by RT-PCR, suggesting that saliva from HGV/GBV-C infected individuals may be infectious (Chen *et al.* 1997). In the other study, HGV/GBV-C was detected at relatively high levels (mean level, 8×10^3 RNA copies/ml) in 13 of 17 saliva samples collected from HGV/GBV-C viraemic volunteer blood donors who were seronegative for HIV, HBV and HCV (Blair *et al.* 1998). However, further studies are required to investigate whether secretion of HGV/GBV-C in saliva may be considered a route of transmission.

1.7.3 HGV/GBV-C AND LIVER TRANSPLANTATION

HGV/GBV-C RNA has been detected in HCV infected and HCV/HBV coinfected liver transplant recipients both before and after operation at frequencies of 15% and 44-53%, respectively (Dickson *et al.* 1997; Fried *et al.* 1997; Haagsma *et al.* 1997; Vargas *et al.* 1997). The increase in the frequency of infection after transplantation has been attributed to blood transfusions received by patients during operation. However, HGV/GBV-C infection did not seem to influence the outcome of liver transplantation in these patients (Dickson *et al.* 1997; Fried *et al.* 1997; Haagsma *et al.* 1997; Vargas *et al.* 1997).

1.7.4 DISEASE ASSOCIATION

The detection of HGV/GBV-C in sera from patients with hepatitis of unknown aetiology suggested that HGV/GBV-C might be responsible for some cases of nonA-E hepatitis (Simons *et al.* 1995a; Linnen *et al.* 1996a). HGV/GBV-C RNA was found in 23% of patients with post-transfusion non-A, non-E hepatitis from the USA and Australia, in 9.5% of patients with chronic non-A, non-C hepatitis (Linnen *et al.* 1996), and in 35% of Italian patients with acute non-A, non-E hepatitis (Fiordalisi *et al.* 1996). In addition, HGV/GBV-C infection was associated with cases of fulminant hepatitis of unknown aetiology in a study where three of six patients with fulminant hepatitis and without evidence of infection with known hepatitis viruses were found to be positive for HGV/GBV-C RNA (Healey *et al.* 1995). It has also been suggested that a

specific strain of HGV/GBV-C might be associated with fulminant hepatic failure since a specific sequence motif within the HGV/GBV-C NS3 region containing six nucleotide mutations was identified in all of eleven German patients with fulminant hepatic failure (Heringlake *et al.* 1996). However, it is difficult to assess the role of HGV/GBV-C in the development of these cases since such patients would have received multiple transfusions before testing (Kao *et al.* 1996a). Another study in Japan showed that HGV/GBV-C RNA was not detected in any of the pre-transfusion sera collected from ten non-A, non-E fulminant hepatitis patients (Kanda *et al.* 1997); after fresh frozen plasma was transfused to these patients, four of them were found to be HGV/GBV-C RNA positive, suggesting that HGV/GBV-C was transmitted by transfusion, and so was not the causative agent.

Other clinical studies have also indicated a lack of causative association between HGV/GBV-C infection and liver disease. The rate of HGV/GBV-C viraemia in blood donors with raised serum ALT levels was reported to be the same as in those with normal ALT levels (Linnen *et al.* 1996; Feucht *et al.* 1997), although another report showed a significant difference (0.8% positive for HGV/GBV-C RNA with normal ALT versus 3.9% with elevated ALT) (Dawson *et al.* 1996). An extensive retrospective study in which the prevalence, clinical features, incidence and natural history of HGV/GBV-C were investigated in a large sample of the Scottish donor population indicated that although there was a relatively high incidence of HGV/GBV-C infection (170-200 new infections per 100,000 donor-years), no evidence for hepatitis or for shared nonhepatic disease

could be identified by clinical examination of the infected donors (Blair *et al.* 1998).

Of 16 HGV/GBV-C positive patients on dialysis with persistent HGV/GBV-C viraemia, none had elevated ALT levels or clinical evidence of liver disease (Masuko *et al.* 1996) and similar observations were made in prospective studies of HGV/GBV-C infected transfusion recipients (Alter *et al.* 1997).

No significant differences have been observed between HGV/GBV-C infected and non-infected patients with chronic hepatitis C virus infection with regard to liver function tests, level of viraemia or biochemical response to interferon (Berg *et al.* 1996; Tanaka *et al.* 1996; Saiz *et al.* 1997; Enomoto *et al.* 1998; Pawlotsky *et al.* 1998). HGV/GBV-C seems to be sensitive to interferon alpha treatment but the response is usually transient, with a relapse after therapy ceases (Karayiannis *et al.* 1997a; Saiz *et al.* 1997a; Pawlotsky *et al.* 1998a).

Because the level of serum ALT does not necessarily correlate with the severity of liver disease, the role of HGV/GBV-C in the pathogenesis of chronic liver disease has also been investigated in various population groups by histological examination (liver biopsy and laparoscopic liver inspection) (Bralet *et al.* 1997; Fried *et al.* 1997; Haagsma *et al.* 1997; Sarrazin *et al.* 1997; Hanley *et al.* 1998). No association between the severity of liver disease and HGV/GBV-C viraemia was identified in a cohort of haemophiliacs previously exposed to non-virus-inactivated coagulation factor concentrates (Hanley *et al.* 1998). A similar conclusion was reached in a study of patients with elevated

aminotransferase levels of unknown aetiology where the histological examination indicated minimal inflammatory activity in both HGV/GBV-C-positive and -negative individuals (Sarrazin *et al.* 1997). In the case of liver transplantation patients, the presence of HGV/GBV-C infection seemed to have a minimal influence on the clinical outcome. Comparison of hepatic histology showed similar degrees of inflammation and fibrosis between HGV/GBV-C-positive and -negative patients whether or not they were coinfected with HCV (Fried *et al.* 1997; Haagsma *et al.* 1997; Pessoa *et al.* 1997).

1.8 QUESTIONS TO BE ANSWERED

Despite extensive epidemiological studies, research into the molecular characterisation of HGV/GBV-C and the clinical significance of HGV/GBV-C infection are at an early stage. At the same time, no definitive conclusions can be made regarding the virus heterogeneity, although some investigations into the sequence diversity of the HGV/GBV-C have been carried out (Muerhoff *et al.* 1996; Muerhoff *et al.* 1997; Wang *et al.* 1997). To address this issue, a detailed analysis was carried out of the phylogenetic HGV/GBV-C groupings obtained by comparison of all currently available complete genome sequences together with examination into the phylogenetic relationships between 5'-UTR sequences from different HGV/GBV-C isolates. Additionally, variation in the HGV/GBV-C 5'-UTR was examined in detail in order to establish its implications for secondary structure and virus classification (Chapter 3. "Identification of

HGV/GBV-C variants by phylogenetic analysis of the 5'-UTR").

Although the genome organization of HGV/GBV-C resembles that of HCV, several interesting questions about the virus structure are still to be answered. The potential lack of a core protein encoded usually at the amino terminus of the polyprotein in flaviviruses together with the limited degree of variation amongst HGV/GBV-C variants suggested by preliminary studies (Erker *et al.* 1996; Linnen *et al.* 1996; Okamoto *et al.* 1997) could have a major effect upon virus pathogenicity and its interaction with the host immune system. Chapter 4 "Buoyant density and sedimentation behaviour of HGV/GBV-C" examines the physicochemical properties of HGV/GBV-C particles by comparison to those of HCV. The presence of the core protein is also investigated by analysing the amino acid sequences of the putative core region of HGV/GBV-C complete genome sequences from different phylogenetic groupings.

At the same time, very little is known concerning the specific features of the terminal regions of HGV/GBV-C genome, which can contain structural elements critical for viral RNA replication, transcription, translation and viral packaging. One way to determine the characteristic structural elements of the HGV/GBV-C 3'-UTR was to perform a comprehensive comparative analysis of the primary sequence and secondary structure of various 3'-UTR sequences. This has also led to the construction of a consensus secondary structure model for the whole 3'-UTR of the HGV/GBV-C genome (Chapter 5. "Analysis of the primary and secondary structure of the HGV/GBV-C").

Nothing is really known about the cellular tropism of HGV/GBV-C. Initially, the virus was considered "hepatotropic" and thought to be associated with human liver disease although the site of HGV/GBV-C replication had not been identified. The absence of a cell system for virus culture *in vitro* represents a major impediment to the investigation of the mechanisms of replication and pathogenicity of the HGV/GBV-C virus. To address this problem, the susceptibility of different cultured cells to HGV/GBV-C infection was tested and analysed (Chapter 6).

Finally, the results of the study described in this thesis may be considered as answers to these crucial questions regarding HGV/GBV-C.

CHAPTER 2

2. MATERIALS AND METHODS

2.1 DENSITY GRADIENT ULTRACENTRIFUGATION

Density gradient ultracentrifugation allows not only the complete separation of several or all the components in a mixture, it also permits analytical measurements to be made.

This method involves a supporting column of fluid whose density increases towards the bottom of the centrifuge tube. The density gradient fluid consists of a suitable low molecular weight solute in a solvent in which the sample can be suspended. Each component will sediment only to the position in the centrifuge tube at which the gradient density is equal to its own density, and there it will remain. In this experiment, sucrose (AnalaR-BDH) was chosen as a density material for preparing the density gradient because of its lack of influence upon the stability of viral particles (Miyamoto *et al.* 1992). A stepwise density gradient was produced ranging from 10% to 60% (w/w) sucrose in TEN buffer (50 mM Tris-HCl pH 8.0, 1 mM EDTA, 150 mM NaCl) as follows: 1 ml of 60% and 0.7 ml each of 50, 40, 30, 20 and 10% sucrose. Each serum sample (0.5 ml) was layered on top of 5 ml of sucrose gradient and centrifuged at 27,800 rpm for 44 h at 10^0C in a Beckman SW50.1 ultracentrifuge rotor. 10 fractions (0.5 ml each) were collected from the bottom of each centrifugation tube and the density of each fraction was determined from its refractive index using an Abbé refractometer. Viral RNA was extracted from 100 μl of each fraction by

Proteinase K method (section 2.2.1). Reverse Transcription (RT) and nested Polymerase Chain Reaction (PCR) were carried out using primers specific for 5'-UTR of HCV and 5'-UTR of HGV/GBV-C, respectively as described in sections 2.2.3 and 2.2.4.

2.2 POLYMERASE CHAIN REACTION

2.2.1 EXTRACTION OF VIRAL RNA FROM SERUM SAMPLES

Viral RNA was extracted by incubating 100 μl of serum sample with 400 μl of lysis buffer (50 mM Tris-HCl pH 8.0, 0.1 M NaCl, 0.5% SDS, 1 mM EDTA, 40 μg/ml polyadenylic acid, 1 mg/ml proteinase K) at 37°C for 90 min. 450 μl of phenol (Rathburn Chemicals Ltd.) was then added to each extraction tube and mixed thoroughly. After centrifugation at 15,000rpm for 10 min, the aqueous layer was transferred to a new eppendorf tube containing 450 μl of chloroform :isoamylalcohol (50:1). This mixture was again vortexed thoroughly and centrifuged at 15,000rpm for 10 min. The aqueous layer was transferred to a new eppendorf tube. The nucleic acid was precipitated overnight at -20°C by the addition of one tenth volume of sodium acetate (pH 5.2) and 2 volumes of ice-cold 100% ethanol (AnalaR) to the extraction tube. The nucleic acid extract was collected by centrifugation at 15,000rpm at 0°C for 15 min. The pellet was then washed with ice-cold 80% ethanol and air-dried at 42°C for 15 min. The dried pellet was resuspended in 25 μl of diethylpyrocarbonate-treated water (DEPC-water). During each extraction process, aliquots of known positive and

negative serum samples (controls) were included in order to ensure that no contamination was present.

The same RNA extraction method was performed to isolate RNA from each supernatant sample obtained from cultured cells at different times of collection and sampling (section 2.5.2).

2.2.2 EXTRACTION OF RNA FROM CULTURED CELLS

The purity and integrity of RNA isolated from tissue or cultured cells are critical for its effective use in application as RT-PCR (sections 2.2.3 and 2.2.4). The RNA was extracted from cell pellets obtained at each time of collection and sampling (section 2.5.2) using RNAgents Total RNA Isolation System (Promega) according to the manufacturer's instructions. This system uses two potent inhibitors of RNase, guanidine thiocyanate and β-mercaptoethanol. In addition, all procedures were performed on ice to significantly slow the rate of RNA degradation. 30 μl of 2 M sodium acetate (pH 4.0) was added to each cell pellet previously homogenized in 300 μl of denaturing solution (26 mM sodium citrate pH 4.0, 0.5% N-laurylsarcosine, 0.125 M β-mercaptoethanol, 4 M guanidine thiocyanate) at different times of collection and sampling (section 2.5.2). The mixture was mixed thoroughly by inverting the tube 4-5 times. 300μl of phenol:chloroform:isoamylalcohol (125:24:1), pH 4.7 was then added to the extraction tubes. After mixing carefully and shaking vigorously for 10 s, the tubes were kept on ice for 15 min. Following this, the tubes were centrifuged at

15,000rpm at 4°C for 20 min and the top aqueous phase was carefully removed to a new tube. The RNA was extracted from the aqueous layer by precipitation with an equal volume of isopropanol overnight at -20°C. The nucleic acid was then collected by centrifugation at 15,000rpm at 4°C for 20 min and washed with ice-cold 80% ethanol. The pellet obtained after a second centrifugation was air-dried and resuspended in 30 μl of nuclease-free water.

2.2.3 REVERSE TRANSCRIPTION OF VIRAL RNA

DNA can be amplified from a single-stranded RNA template by combining a standard PCR protocol with an initial reverse transcriptase reaction. Synthesis of complimentary DNA (cDNA) was carried out at 42°C for 30 min using 5 μl of extracted RNA, 10 units of avian myeloblastosis virus reverse transcriptase AMV (Promega) and 1.5 μM of the outer, antisense primer (used in the corresponding PCR as well) in 20 μl buffer containing 50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 5 mM DTT, 50 mM KCl, 0.05 μg/μl BSA, 15% DMSO, 600 μM each of dGTP, dATP, dTTP, dCTP and 10 units RNAsin (Rnase inhibitor, Promega). The samples were then heated to 98°C for 2 min to heat-inactivate the reverse transcriptase. Prior to transfer to the thermal cycler Techne GeneE for this reaction, each sample was covered with a drop of liquid paraffin to prevent loss of sample mixture because of evaporation.

2.2.4 HOT START PCR AMPLIFICATION

The Polymerase Chain Reaction (PCR) is a technique for the *in vitro* amplification of specific DNA sequences by the simultaneous primer extension of complementary strands of DNA. Problems with PCR sensitivity can arise especially when dealing with low-copy number nucleic acid template. Using nested PCR, the sensitivity and specificity of viral RNA amplification is considerably improved. In this method, the PCR is repeated using a pair of specific primers lying within the sequence within the sequence originally amplified by an external pair of primers. The first PCR was carried out with 5 μl of cDNA (section 2.2.3) and 0.4 units of DNA Taq polymerase (Promega) in 50 μl PCR reaction buffer containing 50 mM KCl, 10 mM Tris-HCl pH 9.0, Triton X-100, 1.5 mM $MgCl_2$, 30 μM each of dGTP, dATP, dTTP, dCTP and 0.25 μM of each of the outer primers (sense and antisense primers, respectively).

The second PCR was carried out in a 20 μl volume for each sample, containing 1 μl of primary PCR product, 8 μM of each of the inner sense and antisense primers, and 0.4 units of DNA Taq polymerase (Promega) in PCR reaction buffer. Each sample was covered with a drop of liquid paraffin before being transferred to Techne GeneE thermal cycler .

The conditions for the first round of PCR were hot start at 80^0C for 2 min to enhance the specificity of the process followed by 30 cycles of 94^0C for 18 s to allow denaturation of cDNA template, 55^0C (for amplification of 5'UTR of HCV and HGV/GBV-C, core region of HCV and NS-3 region of HGV/GBV-

C) or 58°C (for amplification of 3'UTR of HGV/GBV-C) for 21 s to allow primer annealing to single-stranded DNA and 72°C for 90 s to allow strand extension. At the end of the last cycle, samples were heated to 72°C for 5 min to allow termination of incomplete strands.

The second round of PCR was carried out for the same number of cycles on the same thermal cycler at 94°C for 18 s, 55°C for 21 s and 72°C for 90 s. In order to amplify 3'UTR of HGV/GBV-C-for further cloning and sequencing-heminested PCR amplification was performed; the antisense primer was used as outer as well as inner primer. The nucleotide sequences of the primers used for cDNA synthesis and PCR amplification are described in Table 2.1. They were synthesized by Oswell DNA Service, Department of Chemistry, University of Edinburgh.

During each RT-PCR amplification, known positive and negative controls were included in order to ensure that no contamination was present.

2.2.5 STRAND-SPECIFIC PCR AMPLIFICATION

In order to detect whether the replicative intermediates of HCV and HGV/GBV-C were present in the infected cells (section 2.5.2), a sensitive strand-specific nested PCR developed and validated by Mellor et al. (Mellor *et al.* 1998) was performed. The primers for the putative core region of HCV and for the NS3 region of HGV/GBV-C used in this modified nested PCR amplification are listed in Table 2.1. A 'tag' sequence of 20 nucleotides

Table 2.1

Primers used for PCR amplification of HCV and HGV/GBV-C

Primers	Orientation	Primer sequence 5' → 3'	Reference
HCV 5'-UTR			
209	O/AS	ATACTCGAGGTGCACGGTCTACGAGACCT	*
939	O/S	CTGTGAGGAACTACTGTCTT	*
211	I/AS	CACTCTCGAGCACCCTATCAGGCAGT	*
940	I/S	TTCACGCAGAAAGCGTCTAG	*
CORE			
410	O/AS	ATGTACCCCATGAGGTCGGC	#
954	O/S	ACTGCCTGATAGGGTGCTTGCGAG	#
951	I/AS	CAYGTRAGGGTATCGATGAC	#
381	I/S	CAGATCGTTGGYGGAGT	+
380	O/AS-T	**TCATGGTGGCGAATAA**ATGTACCCCATGAGGTCGGC	+,§
379	O/S-T	**TCATGGTGGCGAATAA**CGCCCACAGGACGTYAAGTT	+,§
GBV-C/HGV 5'-UTR			
S4574	O/AS	TGCCACCCGCCCTCACCCGAA	♮
S4571	O/S	AGGTGGTGGATGGGTGAT	♮
S4573	I/AS	GGRGCTGGGTGGCCYCATGCWT	♮
S4572	I/S	TGGTAGGTCGTAAATCCCGGT	♮
5'-UTR/E1			
T1724	O/AS	CCCGCCTGATACAGYGGCCAGCA	this work
T1721	O/S	GGGCAAACGACGCCCACGTACGGTC	this work
T1723	I/AS	GTGCACCCCAGRGCCACMAGGCA	this work
T1722	I/S	TCGCCCTTCAATGYCTCTCTTGRCC	this work
NS3			
3053	O/AS	YTCRTTGATGATGGAACTGTC	♮
3052	O/S	ATCCCCTTTTATGGGCATGG	♮
291	I/AS	TCYTTWCCCCTRTAATAGGC	+
290	I/S	CCCTCGAGCGKATGMGRAC	+
289	O/S-T	**TCATGGTGGCGAATAA**ATCCCCTTTTATGGGCAT	+,§
288	O/AS-T	**TCATGGTGGCGAATAA**YTCRTTGATGATGGAAC	+,§
3'-UTR			
Z3580	O/S	GGTGGTNCATCAATTGGATT	this work
Z3281	I/S	GGTTCTTAGCCCTGCTGATC	this work
Z3282	O&I/AS	AGTAGAACCCGGCCTTTGGG	this work

* Chan et al., 1992.
\# Mellor et al., 1992.
♮ Jarvis et al., 1996.
♮ Mellor et al., 1998.
§ 'Tagged'-primers used in strand-specific PCR (section 2.2.5)
Y = C,T
R = A,G

unrelated to any part of HCV or HGV/GBV-C genomes was present at the 5'end of each outer primer in order to avoid the amplification of falsely primed cDNA products. For this purpose, PCR amplification of a tagged cDNA was performed using only the tag portion of the cDNA primer as one of the primers and an HCV- or HGV/GBV-C-specific oligonucleotide for the opposing primer. For example, to amplify negative strand HCV cDNA, primer 379 (Table 2.1) was used while antisense primer 380 (Table 2.1) was used for positive strand synthesis. In case of HGV/GBV-C, primers 288 and 289 (Table 2.1) were used for positive and negative strand, respectively. In the first round of PCR, the tag sequence alone together with the untagged opposing primer were used (e.g., for HCV negative strand, the tag sequence was the sense primer and 410 was the antisense primer; for HGV/GBV-C negative strand, the tag sequence was the sense primer and 3053 was the antisense one).

The PCR conditions were highly stringent, including pre-denaturation of the template RNA at 70^0C for 10 min prior to addition of AMV reverse transcriptase at 42^0C, heat-inactivation of the enzyme at 98^0C for 10 min after RT reaction followed by HS-PCR (section 2.2.4) to reduce further non-specific priming.

2.2.6 ANALYSIS OF PCR PRODUCTS

The amplified PCR products were visualized under UV light after electrophoresis (at 150 V for 10 min) through a 2% agarose gel in TBE 1X buffer stained with ethidium bromide. Any positive PCR products were detected as fluorescent bands due to the presence of ethidium bromide, an intercalating agent which exhibits a fluorescence under UV light.

2.2.7 LIMITING DILUTION METHOD

Quantification of HCV and HGV/GBV-C RNA was performed using a previously described limiting dilution method in which cDNA was titrated before PCR amplification (Simmonds *et al.* 1990b; Simmonds *et al.* 1990b).

Initial limit dilution was carried out in a series of 10-fold steps which allowed the cDNA to be quantified to within 1 log of its actual concentration. The quantitation was further performed by the addition of a specific cDNA to a number of replicate PCR reactions (5) which would give a Poisson distribution of positive and negative samples and therefore allow the concentration of viral cDNA to be determined. Positive and negative controls were included in each assay and all necessary precautions were taken to avoid contamination. For the RT step, an efficiency of 5% previously established (Zhang *et al.* 1991) was assumed in this assay and the quantification results were expressed as the number of copies of HCV or HGV/GBV-C RNA per ml of serum sample.

2.3 DNA SEQUENCING

Determination of the nucleotide sequence gives a complete description of the amplified DNA, and for many purposes, such as evolutionary and epidemiological studies, it can provide more information than any other analysis method. DNA sequencing is most commonly carried out by strand extension in the presence of dideoxynucleoside triphosphates (ddNTPs) that will terminate synthesis at specific bases (Sanger method). The reactions can be performed directly using amplified DNA as template or cloned PCR products.

2.3.1 PREPARATION OF DNA FOR SEQUENCING

2.3.1.1 SOLID PHASE PURIFICATION OF PCR PRODUCTS

The sequencing reactions are highly sensitive to different components present in the PCR product such as dNTPs, primers and buffer reagents. Therefore, a solid phase purification method (Dynal) was carried out; it allows efficient and simple isolation and purification of target PCR products.

The method is based on the highly specific interaction between streptavidin (protein covalently attached to the surface of supermagnetic, polystyrene beads) and biotin (incorporated into the PCR product). Therefore, a second round of PCR amplification was performed as described in section 2.2.4 using one unlabelled primer and one biotinylated primer (the biotin is attached at its 5'-end in order to maintain 3'-end free for elongation); 100 μl

volume for each sample was prepared containing 1 μl of primary PCR product, 10 μl of PCR reaction buffer, 1 μl of dNTPs (approximately 33 mM), 0.25 μl of biotinylated primer (approx.20 mM), 0.25 μl of unlabelled primer (20 mM), 87.1 μl of pyrogen-free water and 0.4 μl of DNA Taq enzyme (2 units per reaction). The PCR biotinylated products were then purified by using Dynabeads M-280 Streptavidin (Dynal).

For each sample to be sequenced, 20 μl of re-suspended Dynabeads were transferred to a 1.5 ml eppendorf tube and washed twice with BW (Binding and Washing) buffer. The tube was placed in a magnet (Dynal MPC-6) and the supernatant removed. The beads were then resuspended in 40 μl BW, mixed with 40 μl of the amplified biotinylated PCR product and incubated at room temperature for 20 min. During this stage, the beads were resuspended occasionally to allow the immobilization of PCR product to the Dynabeads through the biotin/streptavidin linking system. The beads were then washed in 40 μl of BW and resuspended in 8 μl of 0.15 M NaOH and incubated at room temperature for 10 min; the dsDNA bound to the beads was denatured to ssDNA template. The supernatant containing unbound complementary sense DNA, was then eluted.

The beads were washed once with 50 μl of BW and once with 50 μl of TE to complete the separation of DNA strands. The beads were finally resuspended in 20 μl of TE and stored at 4^0C until sequencing experiment. The presence of Dynabeads in sequencing reactions does not inhibit the enzymatic activity of the sequencing enzyme. During handling, the beads suspension was

mixed by gently pipetting up and down to preserve the integrity of the polystyrene beads.

2.3.1.2 ALKALINE DENATURATION OF PLASMID DNA

The alkaline denaturation method of double-stranded plasmid DNA was carried out in order to perform the sequencing reactions (section 2.3.3).
10 μl of ds plasmid DNA was denatured by adding 1 μl of 2M NaOH/2mM EDTA and incubating at 37^{0}C for 30 min. The mixture was neutralized by adding 1.1 μl of 3 M sodium acetate (pH 4.5-5.5) and the DNA precipitated with 40 μl 100% ethanol at - 20^{0}C overnight.

The denatured DNA was pelleted by centrifugation at 13,000rpm, at 4^{0}C for 10min and then washed with 50 μl of ice-cold 70% ethanol. The washed pellet was redissolved in 6 μl of DEPC-water and used as template for sequencing reactions as described in section 2.3.3.

2.3.2 DIRECT SEQUENCING OF AMPLIFIED PCR PRODUCTS

Direct sequencing of amplified DNA was performed by using Sequenase Version 2.0 DNA sequencing kit (United States Biochemical) with Sequenase™ Version 2.0 T7 DNA polymerase. This enzyme created by *in vitro* genetic manipulation is a genetic variant of bacteriophage T7 DNA polymerase, without a 3' to 5' exonuclease activity. Due to its properties such as high purity, high

processivity, high specific activity and efficient use of nucleotide analogues important for sequencing , Sequenase™ Version 2.0 is a very suitable enzyme to use for chain-termination sequencing.

The chain-termination method involves the synthesis of a DNA strand by Sequenase DNA polymerase *in vitro* using a ssDNA template. This process is carried out in two steps. In the first stage, the purified DNA template is annealed to a oligonucleotide primer and the primer is extended using limiting concentrations of the deoxynucleoside triphosphates (dNTPs), including radioactively labelled dATP. In the second stage, proper mixtures of dNTPs and one of the four 2',3'-dideoxynucleoside 5'-triphosphates (ddNTPs) are used. Since these nucleotide analogues lack the 3'-OH group necessary for DNA chain elongation, enzyme-catalyzed polymerization will end in a fraction of the population of chains at each site where the ddNTP can be incorporated. Due to the radioactively labelled nucleotide included in the synthesis, the labelled chains of different lengths can be visualized by autoradiography after their separation by high-resolution denaturing electrophoresis.

For each template, 2 μl of the appropriate primer, 2 μl of 5X Sequenase reaction buffer, 1 μl dimethylsulphoxide (DMSO) and 6 μl of DNA (bound to magnetic beads) were mixed together in an eppendorf tube. The tube was then heated at 65^{0}C for approximately 5min and allowed to cool slowly to room temperature over a period of 30min. To each annealed template-primer, the following labelling (extension) mixture was added: 1 μl 0.1 M dithiothreitol (DTT), 1 μl of diluted labelling nucleotide mix (1:20), 0.5 μl of [α^{35}-S]dATP

(5 μCi) and 1 μl diluted Sequenase polymerase (1:6 in TE, 2 units).The annealing reaction and labelling mix were mixed throughouly and incubated at room temperature for approximately 5min, during which time the termination reaction was prepared. Four termination mixtures - ddGTP, ddATP, ddTTP, ddCTP -were prepared (80 μM of each dNTP, 8 μM corresponding ddNTP, 50 mM NaCl, 10% DMSO) and 2μl aliqouts of each mix were incubated at 37^{0}C in a waterbath in a microtitre-well plate. When the labelling reaction was complete, 3 μl were transferred to each of the four termination mixes and incubated at 37^{0}C for approximately 5 min and then 4 μl of stop solution were added to each termination reaction.

Following the sequencing reaction, the product was heated to 95^{0}C for approximately 5min and then loaded and electrophoresed on a 5% denaturing polyacrylamide gel (PAGE).

2.3.3 SEQUENCING FROM CLONED DNA

Following the alkaline denaturation step (section 2.3.1.2), the DNA pellet resuspended in 6 μl of DEPC-water was used as a template for sequencing reactions as described above. The primers used in the annealing step were the plasmid primers: 8819 (sense) and 9130 (antisense) for sequencing DNA cloned in pTAg vector (R&D Systems) or pUC_{for} (sense) and pUC_{rev} (antisense) for sequencing DNA cloned in pGemT vector (Promega). The PCR products of HGV/GBV-C 3'-UTR cloned into pGemT vector were sequenced with internal

primer Z4526. The nucleotide sequences of these primers are listed in Table 2.2.

2.3.4 DENATURING PAGE

The quality of gel electrophoresis is an important factor which limits the extent of sequence information that can be determined in a sequencing experiment. Therefore, electrophoresis grade reagents were used and the glass plates were carefully cleaned before pouring the gel by swabbing them with methanol and acetone. The gel solution was prepared by mixing the following reagents : 21 g urea (AnalaR-BDH), 6 ml Ultrapure Sequagel concentrate (50% stock, National Diagnostics), 5 ml 10X Sanger TBE (appendix x), 0.05 g ammonium persulphate and distilled water up to 50 ml. When all the urea was dissolved, 20 μl of TEMED (N,N,N',N'-tetramethylethlediamine, Sigma) was added and the gel solution was poured into the assembly.

Each time, the gels were prepared at least 2 hours prior to use (to allow complete polymerization to occur) and were pre-run for 15-30 min in 1X TBE electrophoresis buffer before loading the samples. The samples were then loaded into the wells of the gel after being heated at 95^0C for approximately 5 min. Gels were run for 2 hours allowing approximately 250 to 300 bases to be read. Gels were then dried and exposed overnight on BIOMAX autoradiography films (Eastman Kodak).

Table 2.2

Primers used for DNA sequencing

Primer	Orientation	Primer sequence 5' → 3'	Comments
B-S4573	I/AS	same as **S4573** with Biotin at 5'	used in direct sequencing (section 2.3.2)
Z-4526	I/S	GGAGGCATGGTGGTTACTAACC	used as internal primer in sequencing DNA clones of 3'-UTR
pTAg primers			
8819	S	GCTATGACCATGATTACGCCAA	used in sequencing from cloned DNA (section 2.3.3)
9130	AS	ACACGTGTGGTCTAGAGC	as above
pGemT primers#			
pUC for	S	GTAAAACGACGGCCAGT	as above
pUC rev	AS	CAGGAAACAGCTATGAC	as above

designed by C. Blake, University of Edinburgh.

2.4 CLONING OF PCR PRODUCTS

2.4.1 USING pTAg VECTOR

Certain thermostable polymerases such as Taq DNA polymerase (Promega) often add single deoxyadenosine (A) overhangs, in a template-independent fashion, to the 3'-ends of the amplified products. These fragments can be easily cloned into the pTAg vector since it is pre-cut and has complementary T overhangs, ready for direct ligation.

pTAg plasmid contains a LacZα peptide sequence, which when functionally produced, complements the N-terminal truncated LacZ peptide synthesized in the competent cells provided. The resulting enzyme, β - galactosidase, cleaves X-Gal (5 Bromo-4 chloro-3 indolyl-β-D galactopyranoside) to give blue colonies. IPTG (isopropyl-thiogalactoside) derepresses the expression of the LacZα gene in cells containing pTAg. When an insert is cloned into the cut pTAg vector the LacZα peptide sequence is interrupted. This interferes with the function of the peptide and usually white colonies are produced on the plates.

2.4.1.1 PREPARATION OF PCR PRODUCT FOR CLONING

The amplified DNA to be cloned was analyzed on 3% agarose gel in 1X TAE buffer. Using long-wave UV light to minimize damage to DNA sample, the band of expected size was excised with a scalpel blade. The DNA was extracted

and purified using Geneclean kit (BIO 101 Inc). The Geneclean kit contains a specially formulated silica matrix called glassmilk that binds single and double stranded DNA without binding DNA contaminants.

The gel slice was transferred in a clean eppendorf tube and 450 μl of NaI stock solution was added. The tube was then placed in a waterbath at 45-55^{0}C for approximately 5 min. When the agarose was completely dissolved, 5 μl of glassmilk suspension was added to the tube. Before using it, the glassmilk tube was vortexed throughouly in order to resuspend the insoluble silica particles. The tube was incubated for approximately 5 min at room temperature to allow binding of the DNA to the silica matrix. The bound DNA was then pelleted in a microcentrifuge for approximately 5s. After removing the supernatant, the DNA pellet was resuspended in 500 μl of ice-cold New wash solution (NaCl/ethanol/water). After that, the pellet was collected in the bottom of the tube by centrifugation and the supernatant was removed. This wash procedure was repeated two more times. Finally, the DNA pellet was resuspended in 50 μl of TE buffer and the tube incubated at 45^{0}C for 3 min. After a brief centrifugation, the supernatant containing the eluted DNA was carefully transferred in a new tube. Eluted DNA was kept at -20^{0}C to be used as insert for cloning.

2.4.1.2 LIGATION OF INSERT USING pTAg VECTOR

Each ligation reaction was set up in 1.5 ml eppendorf tube using 50 ng (1 μl) of pTAg vector, 2 μl of purified PCR product, 1 μl 10X ligase buffer (200 mM Tris-HCl pH 7.6, 50 mM $MgCl_2$), 0.5 μl 10 mM ATP, 0.5 μl (2-3 Weiss units) T4 DNA ligase and the volume made up to 10 μl with nuclease-free water. The solution was mixed gently by pipetting and then incubated overnight at 16^0C. Two control ligations reactions were performed each time to enable accurate interpretation of results: control insert reaction to test the efficiency of the ligation reaction in which 2 μl (5 ng) of control insert provided with the kit was used in place of the amplified fragment; self ligation reaction to test the integrity of the T overhangs of the pTAg vector in which the control insert ligation reaction was carried out without adding the insert (the total volume was made up with nuclease-free water).

2.4.1.3 TRANSFORMATION REACTION

The genotype of the competent cells provided with the cloning kit is the following:*end*A1*hsd*R17(rk_{12}-m_{k12}+)*sup*E44*thi*-1*rec*A1*gyr*A96*rel*A1 *lac*[F'*pro*A^+B^+*lac*1qZΔM15::Tn10(Tc^R)].

The ligation mixture was collected at the bottom of the tube by pulsing in a microfuge and 1 μl of each ligation reaction was transferred into sterile 1.5 ml eppendorf tube on ice. The competent cells were thawed on ice for 5min and

an aliquot of 20 μl was carefully transferred into each tube containing the ligation reaction. After taping gently the tubes to mix the contents, they were incubated on ice for 30min. The tubes were then heat-shocked on waterbath at 42^{0}C for 40 s and returned to ice for 2 min. 80 μl of SOC medium was added to each tube with transformed cells to allow them to grow at 37^{0}C for 1 h in a rotary shaking incubator at approximately 200 rpm.

Luria Broth (LB) agar plates, containing 50 μg/ml Ampicillin, 15 μg/ml tetracycline, 80 μg/ml X-Gal, and IPTG to 0.5 μM, were prepared. Ampicillin is used for selection to ensure the presence of only the transformants with the plasmid that carries the gene for Ampicillin resistance (β-lactamase) while tetracycline is necessary to maintain the F'plasmid containing LacZ gene in the competent cells.

Following the incubation at 37^{0}C , 50 μl of the transformation reaction was spread on each LB selective agar plate. The plates were incubated at 37^{0}C overnight. During transformation and plating, sterile technique was used to avoid contamination.

2.4.1.4 SCREENING OF TRANSFORMANTS FOR DNA INSERTS

Successful cloning of an insert in pTAg vector interrupts the coding sequence of β-galactosidase and white colonies are produced on the plates. White colonies were screened for the presence of DNA insert by colony PCR (Gussow and Clackson, 1990).

Using a sterile toothpick, the edge of a white colony was transferred into 0.5 ml PCR tube containing 20 μl PCR reaction mixture (section 2.2.4) in which plasmid primers 8819 (sense) and 9130 (antisense) were used (Table 2.2). Each solution was covered with a drop of paraffin oil and the PCR tubes placed on Techne Gene E thermal cycler for one cycle of 94^{0}C for 1.5 min, 50^{0}C for 30 s, 72^{0}C for 3 min, and 29 cycles of 94^{0}C for 30 s., 50^{0}C for 30 s, 72^{0}C for 3 min, followed by an incubation at 20^{0}C for 7 min. The samples were then electrophoresed through a 2% agarose/TAE gel containing 0.05 μg/ml EtBr for 20min next to molecular weight markers for reference. Recombinant transformants of correct size were identified and the corresponding colonies from the agar plates were selected.

2.4.1.5 PREPARATION OF PLASMID DNA

The positive colonies identified by PCR as described above were transferred with a sterile toothpick in 3 ml of LB supplemented with 50-100 μg/ml of ampicillin. The cultures were incubated overnight at 37^{0}C in a rotary shaking incubator at approximately 200 rpm.

Next day, 1.5 ml of culture was transferred into an eppendorf tube and spun down in a microcentrifuge for 30 s. The supernatant was discarded and the cell pellet was resuspended in 100 μl GTE (50 mM glucose, 25 mM Tris-HCl pH 8.0, 10 mM EDTA) with lysosyme to degrade the bacterial cell wall. 200 μl of NaOH/SDS solution were added and the tubes were kept on ice for 5 min to

allow the denaturation of DNA. By the addition of 150 μl of solution III (Appendix), the denatured DNA was aggregated and became insoluble. After 5 min, the tubes were spun in a microcentrifuge for 8 min to separate the cell debris and the supernatant containing plasmid DNA was transferred to a new eppendorf tube.

In order to precipitate the DNA, 290 μl of isopropanol were added and the tubes were incubated at room temperature for at least 15 min. The DNA pellet was obtained by centrifuging the tubes at 13,000 rpm for 10min, washed with 500 μl 70% ethanol and the air-dried. The dry DNA pellet was resuspended in 40μl of TE containing RNase to a final concentration of 20 μg/ml and stored at -20^{0}C. Before DNA sequencing, an aliquot of 10 μl was denatured by alkali as described in section 2.3.1.2.

2.4.2 USING pGemT VECTOR

Similar to pTAg vector, pGemT vector (Promega) contains single 3'-T overhangs at the insertion site to prevent recircularization of the plasmid and to provide compatible overhang for the 3'-ends of PCR products generated by certain thermostable polymerases, such as Taq DNA polymerase (Promega). Since it contains a α-peptide coding sequence of the enzyme β-galactosidase, insertional inactivation of the α-peptide allows recombinant clones to be directly identified by colour screening on indicator plates.

The cloning conditions and procedures were similar to those described

for pTAg vector (sections 2.4.1.2 to 2.4.1.5).

JM109 High Efficiency competent cells provided with the pGemT kit were used for transformation reactions. The genotype of JM109 cells is the following : *rec*A1, *end*A1, *gyr*A96, *thi*, *hsd*R17 (r_K^-, m_K^+), *rel*A1, *sup*E44,Δ(*lac-pro*AB), [F', *tra*D36, *pro*AB, *lac*I^q ZΔM15]. Selection for transformants was similar to pTAg vector, on LB/Ampicillin/IPTG/X-Gal.

2.5 CELL CULTURE

2.5.1 CELL PREPARATION

The following cultured cells were used for *in vitro* infection experiments: peripheral blood mononuclear cells (PBMCs), U937 cells, and HepG2 cells. For each culture, cell counts and viability were assessed routinely by trypan blue exclusion test (section 2.5.3). Sterile techniques were carried out at all stages during the experiments to avoid contamination.

PBMCs were isolated from fresh heparinized blood of healthy donors by centrifugation on Ficoll-Hypaque gradients (Lymophoprep, Nycomed) by C. Forbes (University of Edinburgh). The cell suspension was then washed in PBS and resuspended in RPMI 1640 growth medium (Gibco) supplemented with 10% heat-inactivated fetal calf serum (FCS, Gibco), 100mM L-Glutamine (L-Glu, Gibco), antibiotics (penicillin/streptomycin) at concentration of 10^4IU/ml (Gibco), and interleukin-2 (IL-2, NIBSC) at concentration of 10^4IU/ml. The cell

suspensions were maintained at 37^{0}C in a 5% CO_2 atmosphere (ICN flow incubator). Prior to virus inoculation, the PBMCs were incubated with phytohaemagglutinin (PHA, Murex) at concentration of 5μg/ml for 48h to stimulate cell division.

The U937 continuous macrophage-like cell line derives from a human histiocytic lymphoma and exhibits morphologic and functional characteristics of normal human macrophages (Sundstrom and Nilsson, 1976). Unlike peripheral macrophages, U937 cells are nonadherent. The cells were maintained in suspension in RPMI 1640 medium containing 10% FCS, L-Glu and antibiotics. U937 cells were kept at 37^{0}C in a 5% CO_2 atmosphere and passaged weekly at a ratio of 1:4.

Differentiation of U937 cells to macrophages was achieved by addition of phorbol 12-myristate, 13-acetate (PMA) to cultured cells. Cell suspension was washed once with PBS by centrifugation at 1000rpm for 5min and resuspended in fresh culture medium (supplemented RPMI 1640) with PMA (Sigma) at a final concentration of 160 nM. The cells were then aliquoted into 6-well plates (Nunc) at a concentration of 6 x 10^5 cells/ml and incubated for 24h at 37^{0}C in a 5% CO_2 atmosphere. Prior to virus addition, the U937 cell monolayers were washed gently to remove unattached cells and fresh culture medium without PMA was added to the wells.

HepG2 cells are well differentiated hepatoblastoma cells derived from human liver tumour, with biosynthetic characteristics similar to human hepatocytes. Initially, the cells were cultured in RPMI 1640 medium

supplemented with 10% FCS and antibiotics (10^4IU/ml) as monolayers in culture flasks (Nunc) at 37^0C in a 5% CO_2 atmosphere. The culture medium was renewed twice a week. Before the *in vitro* infection experiments, confluent HepG2 cell monolayers were harvested by trypsinization (section 2.5.3) and the cells were washed twice and resuspended at 1 x 10^6 cells/ml in FCS-free Ham's F12 medium supplemented with growth factors, antibiotics and other supplements (Appendix). At the same time, in order to upregulate the low-density lipoprotein (LDL) receptors on the cell surface, lovastatin (Merck, Sharp & Dohme) was added to Ham's F12 medium to a final concentration of 10 μM. The cells were allowed to settle in 6-well plates at 37^0C in a 5% CO_2 atmosphere. After 72h-exposure to lovastatin, the medium was removed, and the cell monolayer washed three times with PBS. Fresh serum-free Ham's F12 supplemented medium (2 mL) was then added to each well.

2.5.2 INOCULATION PROCEDURES

HCV and HGV/GBV-C positive serum samples together with a negative control of a HCV and HGV/GBV-C negative serum were used as inocula for *in vitro* infection of the cells (Table 6.1). The viral titres of positive samples (number of RNA copies/ml serum) were determined using the limiting dilution method (section 2.2.7).

IN VITRO INFECTION OF PBMCs

After PBMCs stimulation by PHA (section 2.5.1), the cells were washed three times with PBS and resuspended in fresh RPMI 1640 medium with supplements. Cells were then aliquoted into 6-well plates at a concentration of 1 x 10^6 cells/ml and inoculated with three separate serum samples (10 μl inoculum) for 4h (4h-incubation experiment) or 24h (24h-incubation experiment) at 37^0C in a 5% CO_2 atmosphere. After incubation, the cells were washed three times with PBS to remove inoculum and then maintained in culture flasks with RPMI 1640 medium containing 10% FCS, antibiotics, L-Glu and IL-2. Cultures were harvested on days 1, 3, 8 post-inoculation (p.i.) for the 4h-incubation experiment, and on days 1, 2, 7, 10 p.i. for the 24h-incubation experiment. At each collection point, 1 ml aliquots of cell suspensions were removed and fresh medium was added. Following the centrifugation of cell aliquots at 1000 rpm for 5 min, the supernatants were kept at -40^0C for RNA extraction (section 2.2.1). The corresponding cell pellets were washed three times with PBS, resuspended in 300 μl of denaturing solution (Promega) and also stored at -40^0C for RNA extraction (section 2.2.2).

IN VITRO INFECTION OF U937 CELLS

Separate experiments using undifferentiated U937 cells (cell suspension) and differentiated (PMA-treated) U937 cells (monolayers) were carried out in

order to investigate the susceptibility of these cells to infection by HCV and HGV/GBV-C.

Using U937 cell suspension (section 2.5.1), cells were washed once with PBS, adjusted to a density of 2 x 10^5 cells/ml in fresh RPMI 1640 medium with 10% FCS, L-Glu and antibiotics in Bijou bottles (Sterilin) and incubated with 25 μl inoculum for 4h at 37^0C in a 5% CO_2 atmosphere. After incubation, the cells were pelleted by centrifugation at 1000 rpm for 5 min, washed three times with PBS, and the cultures were maintained in flasks at a density of 2 x 10^5 cells/ml in fresh medium. They were harvested at days 1, 2, 3 and 4 p.i. and the samples (supernatants and cells) stored at -40^0C for RNA extraction in the same way as described for PBMCs cultures.

A second experiment using U937 cell suspensions inoculated with the same inocula (25 μl or 100 μl per each culture) for 24h was performed and maintained as described above. In this case, the inoculated cultures were kept for a longer period (18 days) and they were harvested at days 3, 9, 15 and 18 p.i.

In vitro infection of differentiated U937 cells were carried out in 6-well plates by incubation of cell monolayers with 100 μl inoculum in 1.5 ml medium for 4h. Following exposure, cells were washed gently three times with PBS to remove inoculum and fresh medium (RPMI 1640 + 10%FCS + L-Glu + antibiotics) was added to each well. At days 1, 2, 3 and 4 p.i., cell monolayers and medium were removed from the corresponding wells mechanically using disposable cell scrapers (Nunc) and pelleted by centrifugation at 100 rpm for 5 min. Supernatants were kept at -40^0C for RNA extraction together with the cell

pellets resuspended in 300 μl denaturing solution after being washed three times with PBS.

IN VITRO INFECTION OF HepG2 CELLS

The HepG2 cell monolayers prepared as described above (section 2.5.1) were incubated with 100 μl inoculum for 90 min at 37^{0}C in a 5% CO_2 atmosphere. After incubation, the monolayers were washed three times with PBS and grown in serum-free highly supplemented Ham's F12 medium. For RNA extraction, at days 3, 7, 14 and 21 p.i., the culture media were centrifuged at 1000rpm for 5 min and supernatants stored at -40^{0}C; the pelleted cells were washed three times with PBS and then resuspended in 300 μl denaturing solution to be kept at -40^{0}C.

2.5.3 TRYPAN BLUE EXCLUSION TEST

During cell culture experiments, the stage of growth for each cell culture was routinely estimated from its appearance under the microscope and counting the cells and assessing their viability.

Cell counting was performed by trypan blue exclusion test in which the cells mixed with viability dye are placed in an optically flat chamber (improved Neubauer haemocytometer) under the microscope. For this purpose, 0.1 ml cell suspension was mixed with 0.9 ml trypan blue. A small drop from this mixture

was carefully transferred with a micropipette under the coverslip of improved Neubauer haemocytometer. After placing the slide under the microscope, the unstained (living) cells lying in 4 large corner squares of the grid were counted. To calculate the number of cells/ml of initial suspension, the mean value/corner square was multiplied by 10^4 and then by dilution (1:10). For example:

$(35 + 40 + 42 + 36)/4 \times 10^4 \times 10 = 38 \times 10^5 = 3.8 \times 10^6$ cells/ml culture.

To obtain a count statistically representative of the cell population, two counts/cell culture have been carried out each time.

In case of HepG2 cell monolayers maintained initially in RPMI 1640 medium with 10% FCS and antibiotics (section 2.5.1), the cells were trypsinized in order to be counted and seeded in FCS-free highly supplemented Ham's F12 medium. For this purpose, the culture medium was discarded and monolayer washed twice with 5 ml PBS. 1 ml 0.25% trypsin (Gibco BRL) in PBS was added and run few times over monolayer. When the cell layer became opaque, the trypsin was discarded and the culture flask incubated at 37^0C for 3-5 min.

2.6 PHYLOGENY ANALYSIS

Data generated as a result of sequencing nucleic acids and proteins can be used to study molecular phylogeny, the reconstruction of the evolutionary history of genes or organisms. The evolutionary relationships amongst genes or groups or organisms can be represented graphically as phylogenetic trees. These trees or dendodragms consist of nodes and branches that connect them. The

branching pattern of a tree is called topology and the nodes are either external, at the end of a branch, or internal, where two branches form a fork. Generally, the characters at the external nodes are the data points that are used to construct the tree (group of organisms, a gene, a region of a gene, etc.) and are called operational taxonomic units (OTUs). Phylogenetic trees can be either rooted or unrooted. A rooted tree indicates the direction of evolution, the root representing the common ancestor of all the OTUs studied. An unrooted tree indicates the relationship amongst the OTUs but does not define the evolutionary path. An unrooted tree can become rooted by adding an OTU which is evolutionarily related to the OTUs studied and had diverged from the other OTUs prior to their divergence from one another.

There are different methods available for constructing phylogenetic trees from molecular data (Nei, 1987): distance methods and discrete-character methods. In distance methods, evolutionary distances are calculated for all OTUs which are studied and the phylogenetic tree is constructed by certain principles and algorithms. Discrete-character methods involve the use of discrete character states (nucleotides or amino acids at a site), and a tree is constructed by considering the evolutionary relationships of OTUs at each character state.

In general, in the construction of phylogenetic trees, the principle of minimum evolution is often used. The standard algorithm of the tree-making methods based on this principle is to examine all possible topologies or a certain number of topologies that are likely to be close to the true tree and to select one which shows the smallest amount of total evolutionary change as the final

tree. Computer simulations have shown that the neighbor-joining (NJ) method (Saitou and Nei, 1987), which has been used in this study as part of Molecular Evolutionary Genetics Analysis (MEGA) package (Kumar *et al.* 1993), is one of the most efficient methods in obtaining the correct tree topology under the principle of minimum evolution. According to NJ method, a pair of neighbors is a pair of OTUs connected through a single interior node in an unrooted, bifurcating tree. it is possible to define the topology of a tree by successively joining pairs of neighbors and producing new pairs of neighbors. Since this method produces an unrooted tree, it usually requires an outgroup OTU to find a root. In MEGA, unless outgroup OTUs are specific, the root is sometimes given at the midpoint of the longest route connecting two OTU's in the tree under the assumption of a constant rate of evolution.

Testing the reliability of a tree obtained can be done using two different methods: the maximum likelihood method (Felsenstein, 1993) and the bootstrap test. The maximum likelihood method examines the reliability of every interior branch of the tree. The bootstrap test involves randomly resampling the data from which the tree was constructed, producing a new tree with the resampled data. This process is repeated several times, and the frequency at which particular branches are observed in the newly constructed tree is calculated giving a probability to each branch in the original tree. MEGA package provides bootstrap test for NJ trees.

Another software package used in sequence analysis was Simmonic 2000 sequence editor program created by Dr. P.Simmonds (University of Edinburgh).

This program proved to be very useful because it provides a convenient format to store aligned sets of sequences generated by DNA sequencing and from import of published sequences from other databases such as GenBank. The format was specially designed to hold sequence data and to provide a particular sequence environment for sequence editing including information on how the sequences are numbered, whether they are translated (and reading frame) or whether they should be viewed in reverse complement. Simmonic 2000 sequence editor program provides, through its commands, the interface to a number of sequence analysis packages such as PHYLIP and MEGA. For example, its file operations allow loading, saving and exporting sequences to and from Simmonic file format to external files either for storage or for further analysis.

2.7 PREDICTION OF RNA SECONDARY STRUCTURE

The secondary structure of RNA is important for many functions, including regulation of transcription, translation and replication. There are two methods of predicting secondary structure: phylogeny and energy minimization. Phylogeny relies on alignment and subsequent folding of several sequences into similar structures for functionally analogous RNA. Energy minimization relies on thermodynamic parameters and computer algorithms to determine the minimum and near minimum free energy foldings of an RNA molecule.

A very useful program in predicting RNA secondary structure in this study was RNAdraw program (Matzura and Wennborg, 1996). The optimal structure/basepair-probability matrix/heat curve calculation algorithms in this program were imported from the Vienna RNA package (Hofacker *et al.* 1994). In this program, G-U base pairs were accepted for the analysis, in addition to Watson-Crick base pairs.

CHAPTER 3

3. IDENTIFICATION OF HGV/GBV-C VARIANTS BY PHYLOGENETIC ANALYSIS OF THE 5'-UNTRANSLATED REGION

3.1 INTRODUCTION

Sequence diversity across the HGV/GBV-C genome is significantly lower than that observed with HCV isolates. While major HCV genotypes vary by 31-35% across the complete genome (Simmonds, 1995), the diversity between HGV/GBV-C isolates was reported to be only 14-16% (Erker *et al.* 1996; Mukaide *et al.* 1997; Okamoto *et al.* 1997; Wang *et al.* 1997).

Three genotypes of HGV/GBV-C have been defined by phylogenetic analysis of full-genomic sequences of different HGV/GBV-C isolates (Erker *et al.* 1996; Nakao *et al.* 1997; Okamoto *et al.* 1997; Katayama *et al.* 1998) while similar analysis of nine complete genome sequences suggested that there is one single genotype of HGV/GBV-C (Wang *et al.* 1997). At the same time, in assessing HGV/GBV-C variability, sequence analysis of different subgenomic regions had indicated conflicting results. For example, phylogenetic analysis of 5'-terminal 600 nucleotide-sequences amplified from the sera of HGV/GBV-C infected individuals from around the world indicated the existence of the same three major groups or types (1, 2, 3) as well as several subgroups or subtypes (1a, 1b, 2a, 2b) (Muerhoff *et al.* 1996b). The distribution of the HGV/GBV-C groups is well correlated with the geographic origin of the variants. Group 1 includes the GBV-C prototype and African isolates, group 2 consists of isolates

mainly from Europe and North America while group 3 is related to isolates from Asian countries. Similar groupings were observed from comparisons of fragments of the 5'-UTR (Fukushi *et al.* 1996; Orito *et al.* 1996; An *et al.* 1997; Katayama *et al.* 1997; Mukaide *et al.* 1997).

In contrast, sequence analysis of different subgenomic coding regions did not reveal the same HGV/GBV-C phylogenetic groups. For example, comparison of sequences of 100 or 199 nucleotides in the NS3 region could not differentiate between Indonesian, American and Asian HGV/GBV-C isolates (Tsuda *et al.* 1996) or between European, American, Asian and African variants, respectively (Pickering *et al.* 1997a). Another study based on phylogenetic analysis of 446 nucleotide-fragments of NS3 region from Chinese isolates suggested that there is only one predominant or major HGV/GBV-C genotype in the world (Wang and Jin, 1997). On the other hand, when HGV/GBV-C isolates from Taiwan were compared with isolates from Africa and North America using a fragment of 118 nucleotides of NS3 domain, it was suggested that a genetic heterogeneity exists amongst isolates from different geographic areas (Kao *et al.* 1996b).

When the primary sequence of HGV/GBV-C isolates from Japan corresponding to a longer fragment of 594 nucleotides of NS3 region was determined and compared with that of a 385 nucleotide-fragment of the 5'-UTR of the same isolates, the existence of three major clusters associated with the geographic origin of the variants was confirmed (Katayama *et al.* 1997). However, the groupings were more clearly distinguished by comparison of the

5'-UTR sequences than by comparison of NS3 sequences.

No groupings were observed after analysis of NS5B sequences as comparative analysis of a 354 nucleotide-fragment indicated that the HGV/GBV-C isolates collected from different countries (USA, Germany, Russia, Brazil, Spain, Indonesia, Zaïre, Vietnam) belong to the same type (Viazov *et al.* 1997).

These observations are very different from similar studies of HCV variability. Comparison of nucleotide sequences of complete genomes or subgenomic regions resulted in the classification of at least six major HCV genotypes, some of which being further divided in subtypes (Simmonds *et al.* 1993; Simmonds *et al.* 1994). Relatively short sequences (e.g. 222 nucleotide-fragment of NS5) provided effective discrimination of types, subtypes and isolates (Simmonds *et al.* 1993a) while sequence comparisons of the core, NS3, NS5 and E1 regions indicated equivalent relationships between HCV variants (Chan *et al.* 1992; Simmonds *et al.* 1994).

All of the six major HCV genotypes can be now identified from analysis of any coding region of HCV genome (Simmonds *et al.* 1994) and may also be imperfectly deduced from the presence of particular sequences in the 5'-UTR (Simmonds *et al.* 1993b). Although highly conserved, this region contains several type-specific nucleotide polymorphisms that would allow the identification of genotypes 1 to 6 (Simmonds *et al.* 1993b). However, identifying novel HCV genotypes by sequence comparison only in this region proved unreliable. For example, new HCV variants from Vietnam and Thailand had sequences in the 5'-UTR similar or identical to those of type 1, while the phylogenetic analysis

in the NS5 and E1 regions revealed that they were distantly related to type 6a genotype (type 6 group) (Mellor *et al.* 1995). As a result, a genotyping assay in which sequences in the core region of HCV were amplified and cleaved by restriction enzymes was developed to differentiate type 1 variants from type 6 variants (Mellor *et al.* 1996).

In the face of the contradictory findings regarding HGV/GBV-C heterogeneity, a detailed examination was undertaken of the phylogenetic groupings obtained by comparison of all currently available complete genome sequences and to determine if a specific region within 5'-UTR can be identified as optimal for distinguishing between these HGV/GBV-C groupings.

3.2 HGV/GBV-C ISOLATES USED IN THIS ANALYSIS

Thirty-one complete genome sequences of HGV/GBV-C isolates were available in the GenBank database as of May 1998. However, five of these represented pairs or triplets of epidemiologically linked isolates that had over 99% nucleotide identity with each other: GSI85 (D87262) and GSI93 isolates (D87263); T55875 (AF031827), HGV-1217 (AF031828) and HGV-1539 (AF031829); and K3732 (AB008335) and K1737 (D87709). Only one of each of these groups of sequences were included leaving a total of twenty-six sequences. K10-HGV isolate, which sequence is incomplete (6546 nucleotides), was also included in the analysis (Table 3.1).

Table 3.1 HGV/GBV-C complete genome sequences used in this analysis

Isolate	GenBank accession number	Geographical origin	Length (nucleotides)
1. GT110	D90600	Japan	9395
2. PNF2161	U44402	USA	9392
3. HGV-Iw	D87255	Japan	9375
4. T55875	AF031827	Japan	9351
5. CG01BD	AB003289	Japan	8953
6. R10291	U45966	USA	9103
7. GBVC-EA	U63715	East Africa	9367
8. K2141	D87713	Japan	9391
9. K1741	D87710	Japan	9391
10. K606	D87708	Japan	9256
11. K1789	D87711	Japan	9391
12. GSI85	D87262	Japan	9391
13. GT230	D90601	Japan	9390
14. HGV-IFM1	AB008342	Japan	9387
15. K1775	D87715	Japan	9132
16. K1916	D87712	Japan	9391
17. K1668	D87714	Japan	9391
18. HGVCN	U94695	China	9213
19. CG07BD	AB003290	Japan	8952

Table 3.1 continued

20. G13HC	AB003293	Japan	8930
21. BG1HC	AB003288	Japan	8930
22. AF006500	AF006500	China	8982
23. GBV-C	U36380	West Africa	9377
24. CG12LC	AB003291	Japan	8965
25. G05BD	AB003292	Japan	8930
26. HGVC 964	U94695	China	9128
27. K10-HGV	*	Thailand	6546

* Dr. S. Songsivlai (personal communication)

Additional 5'-UTR sequences were obtained through analysis of RNA extracted from serum or plasma samples obtained from 14 HGV/GBV-C infected patients with chronic hepatitis from Pakistan, from nine pregnant women coinfected with human immunodeficiency virus from Zaïre and from 12 Edinburgh haemophiliacs. RNA from Pakistan and Zaïre samples was amplified by RT-PCR (sections 2.2.3-2.2.4) using 2 sets of primers derived from the HGV/GBV-C 5'-UTR: S4571 (sense outer), S4572 (sense inner), S4573 (antisense inner) and S4574 (antisense outer) (Table 3.2). The PCR products were re-amplified from primary products using primers S4572 and biotinylated S4573 and sequenced directly from magnetically separated single strands after immobilization on streptavidin-coated beads (sections 2.3.1.1 and 2.3.2). RNA from haemophiliac samples was amplified with primers T1721 (sense outer), T1722 (sense inner), T1723 (antisense inner) and T1724 (antisense outer). The PCR products were purified and cloned using pTAg vector (section 2.4.1) and sequenced using both sense and antisense plasmid primers and T7 DNA polymerase (section 2.3.3). The consensus sequence of one to five clones for each sample was used for phylogenetic analysis.

Nucleotide positions within the HGV/GBV-C genome are given relative to the AUG codon that is likely to represent the initiation codon for translation of the GBV-C polyprotein (Simons *et al.* 1996). Therefore, the positions in the 5'-UTR are given negative numbers relative to this codon.

Table 3.2 HGV/GBV-C 5'-UTR sequences obtained for this study

Geographical origin of samples	Primers* used in nested PCR	DNA sequencing
Edinburgh (n=12)	T1721-T1724 T1722-T1723	pTAg primers (8819, 9130)
Pakistan (n=14)	S4571-S4574 S4572-S4573	B-S4573
Zaire# (n=9)	S4571-S4574 S4572-S4573	B-S4573

B-S4573 = biotinylated S4573 primer.

* = primer sequences are listed in Table 2.1 (PCR) and Table 2.2 (DNA sequencing).
#= sequences determined by Fiona Davidson, BTS-Edinburgh.

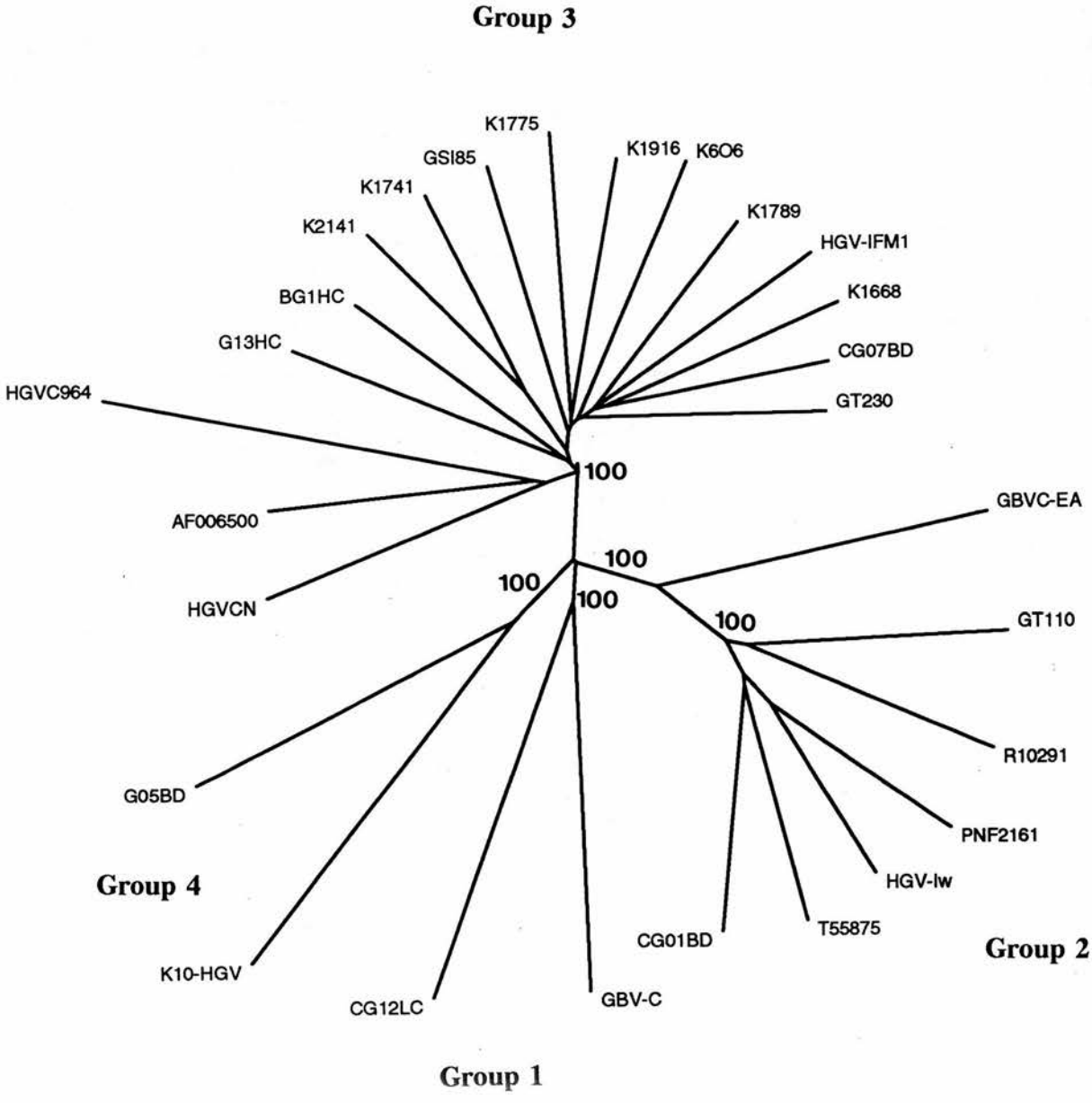
3.3 RESULTS

COMPLETE GENOME SEQUENCE ANALYSIS

The complete genome sequences of twenty-six HGV/GBV-C isolates were manually aligned using the Simmonic 2000 sequence editor program and evolutionary distances determined using the DNADIST program (maximum likelihood parameter) of the Phylip package 3.5c (Felsenstein, 1993). A consensus phylogenetic tree was constructed using the CONSENSE program and the neighbor-joining method of the NEIGHBOR program for 100 bootstrap replicating datasets generated using the SEQBOOT program. The resulting unrooted phylogenetic tree was displayed using TREEVIEW 1.5.2 (Page,R.M.D., 1998). Bootstrap values over 70% were considered as providing evidence for the observed phylogenetic groupings.

This phylogenetic analysis revealed the presence of four independent clusters, each of which was supported by bootstrap resampling (Figure 3.1). For example, sequences CG01BD, T55875, HGV-Iw, GT110 and GBVC-EA grouped with PNF2161 and R10291 isolates in 100% of bootstrap replicates, and together comprise group 2 or G2 (Nakao *et al.* 1997; Okamoto *et al.* 1997; Katayama *et al.* 1998). Another cluster was represented by sequences grouping with GT230 and GSI85 (group 3 or G3) (Nakao *et al.* 1997; Okamoto *et al.* 1997). The prototype isolate, GBV-C, segregated with CG12LC sequence as group 1 in 100% of bootstrap replicates. The G05BD and K10-HGV sequences represent

Figure 3.1 Phylogenetic tree of HGV/GBV-C complete genome sequences. Evolutionary distances between sequences were calculated using Phylip DNADIST (maximum likelihood distances). The tree was constructed using Phylip NEIGHBOR program for 100 bootstrap replicates and displayed using TREEVIEW 1.5.2 (Page,R.M.D., 1998).



an additional group also with 100% bootstrap support. Although incomplete (6546 nucleotides), the K10-HGV sequence was considered in this analysis because its presence did not modify the pattern of groupings; the same clusters were obtained when the same phylogenetic analysis was carried out without K10-HGV isolate (data not shown).

Next, phylogenetic analysis of the same HGV/GBV-C isolates was carried out for only the 5'-UTR sequences in order to test if the variants would form the same clusters as obtained by analysis of the complete genome sequences. For this purpose, the nucleotide sequences corresponding to the 5'-UTR of the HGV/GBV-C isolates were manually aligned using the Simmonic 2000 sequence editor program and analyzed using the MEGA package (Kumar *et al.* 1993) (Figure 3.2). Phylogenetic trees were generated using the neighbor-joining algorithm for 100 bootstrap replicates. Comparison of the entire 5'-UTR sequences (from position -553 to position -1) provided bootstrap support of over 81% for groups 1, 2 and 3 but group 4 was observed in only 38% of bootstrap replicates (Figure 3.3). Surprisingly, the degree of bootstrap support for groups 1-4 was very similar when progressively shorter segments of the 5'-UTR were analysed by removing up to 200 nucleotides from either the 5'-end (Figure 3.3) or from the 3'-end (Figure 3.4). The level of bootstrap support became low (51% for group 2 and 33% for group 3) when 250 nucleotides were removed from the 5'-end (Figure 3.3). Similar result was obtained by analyzing fragments where 250 or 300 nucleotides were deleted from the 3'-end, respectively (the bootstrap support for group 3 was less than 70%) (Figure 3.4). This phylogenetic

Figure 3.2 Sequence alignment of the 5'-UTR of HGV/GBV-C complete genome sequences. Sequence identity with GT110 is indicated by dots and the missing information by spaces. Nucleotide positions are numbered relative to AUG codon at the beginning of the E1 protein of the prototype isolate GBV-C (accession number U36380). The group number is indicated at the right.

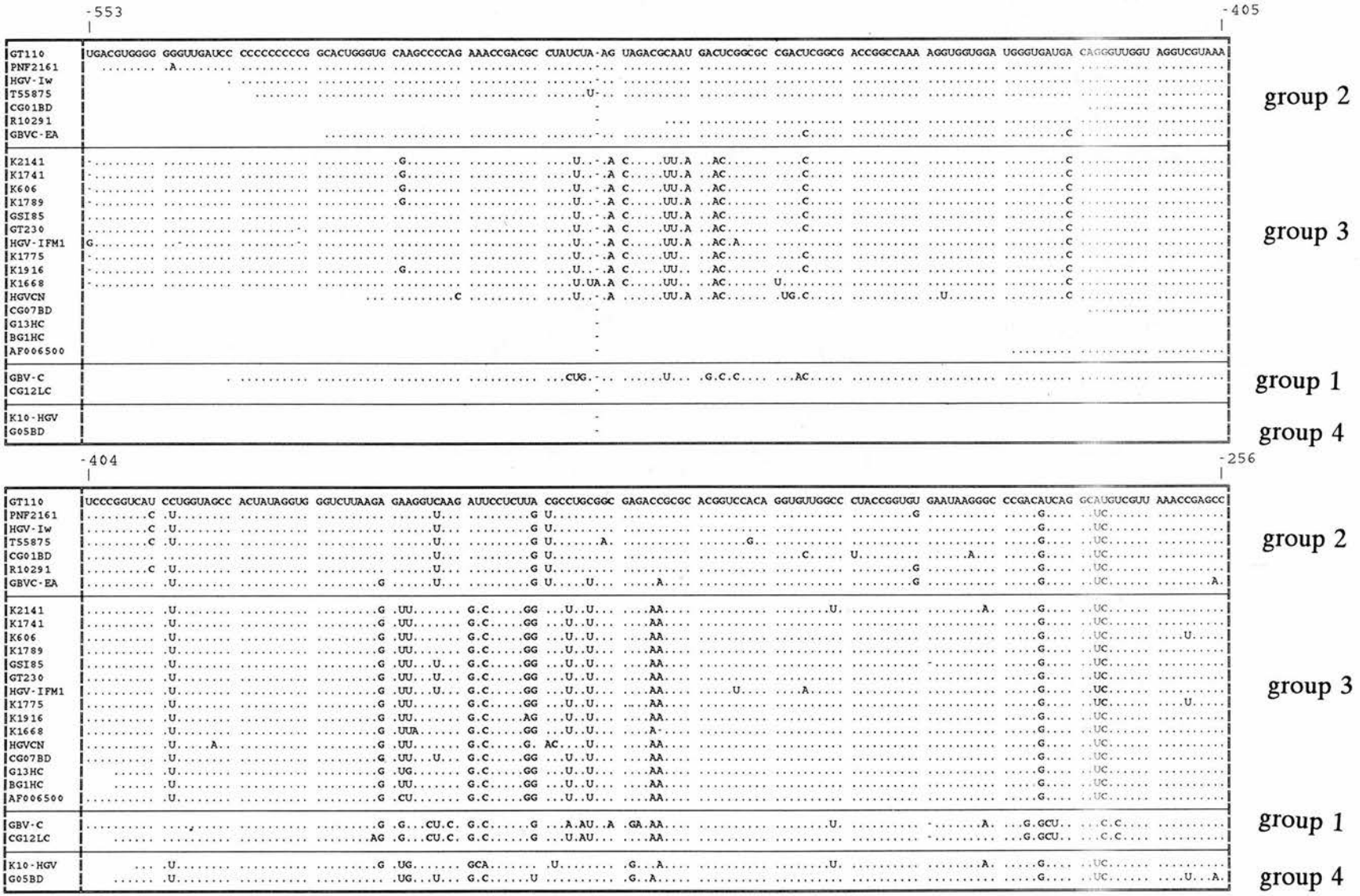

-553
-405
GT110 UGACGUGGGG GGGUUGAUCC CCCCCCCCG GCACUGGGUG CAAGCCCCAG AAACCGACGC CUAUCUA-AG UAGACGCAAU GACUCGGCGC CGACUCGGCG ACCGGCCAAA AGGUGGUGGA UGGGUGAUGA CAGGGUUGGU AGGUCGUAAA
PNF2161
HGV-Iw
T55875
CG01BD
R10291
GBVC-EA
group 2
K2141
K1741
K606
K1789
GSI85
GT230
HGV-IFM1
K1775
K1916
K1668
HGVCN
CG07BD
G13HC
BG1HC
AF006500
group 3
GBV-C
CG12LC
group 1
K10-HGV
G05BD
group 4
-404
-256
GT110 UCCCGGUCAU CCUGGUAGCC ACUAUAGGUG GGUCUUAAGA GAAGGUCAAG AUUCCUCUUA CGCCUGCGGC GAGACCGCGC ACGGUCCACA GGUGUUGGCC CUACCGGUGU GAAUAAGGGC CCGACAUCAG GCAUGUCGUU AAACCGAGCC
PNF2161
HGV-Iw
T55875
CG01BD
R10291
GBVC-EA
group 2
K2141
K1741
K606
K1789
GSI85
GT230
HGV-IFM1
K1775
K1916
K1668
HGVCN
CG07BD
G13HC
BG1HC
AF006500
group 3
GBV-C
CG12LC
group 1
K10-HGV
G05BD
group 4

Figure 3.2 continued

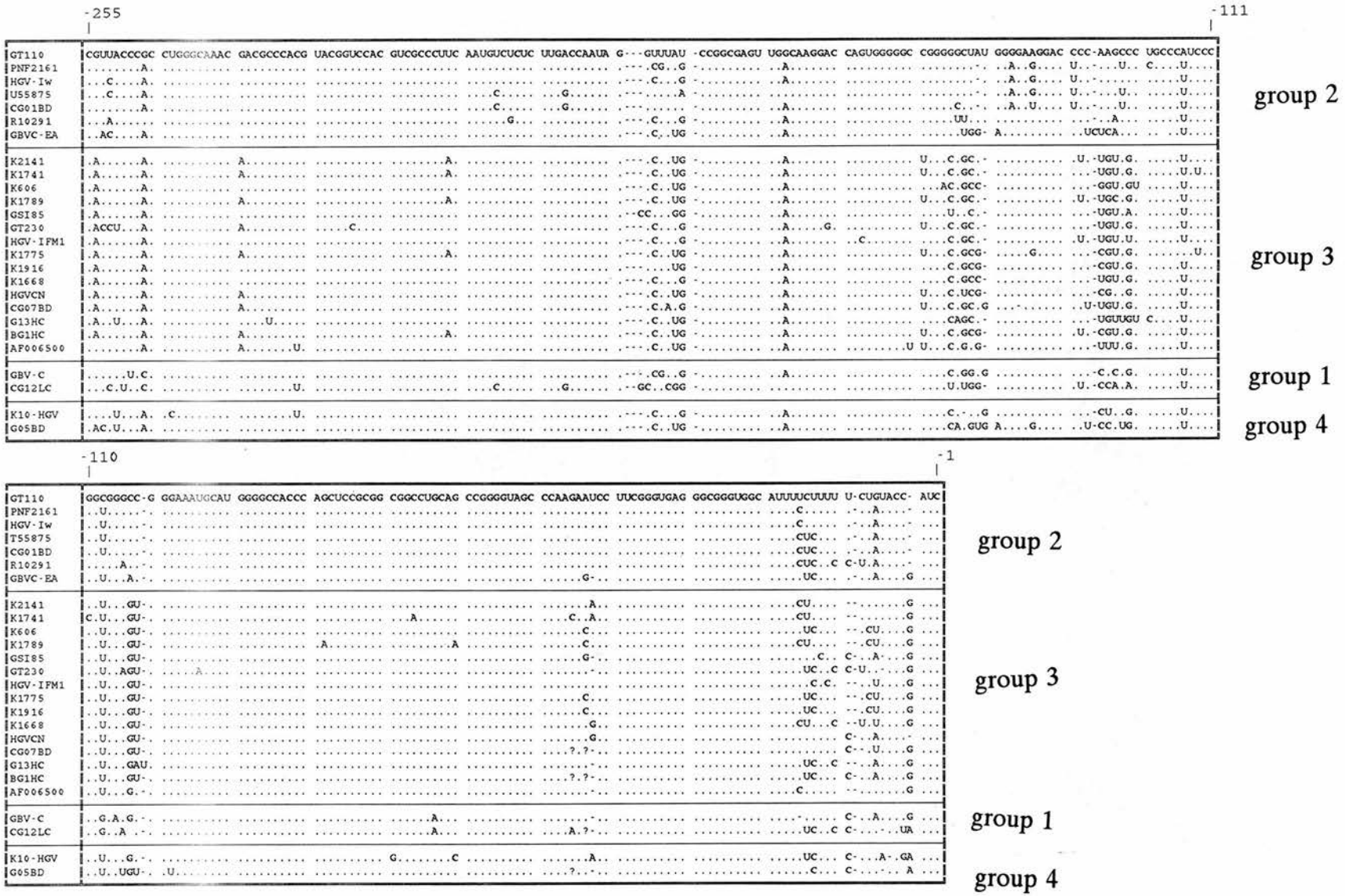


Figure 3.3 Bootstrap support for HGV/GBV-C phylogenetic groups based on the analysis of fragments of the 5'-UTR. Phylogenetic analysis was performed on progressively smaller fragments obtained by removal of nucleotides from the 5'-end of the HGV/GBV-C 5'-UTR using the MEGA package (p-distance) for 100 bootstrap replicates.The level of bootstrap support for each phylogenetic group was plotted against the number of nucleotides removed.

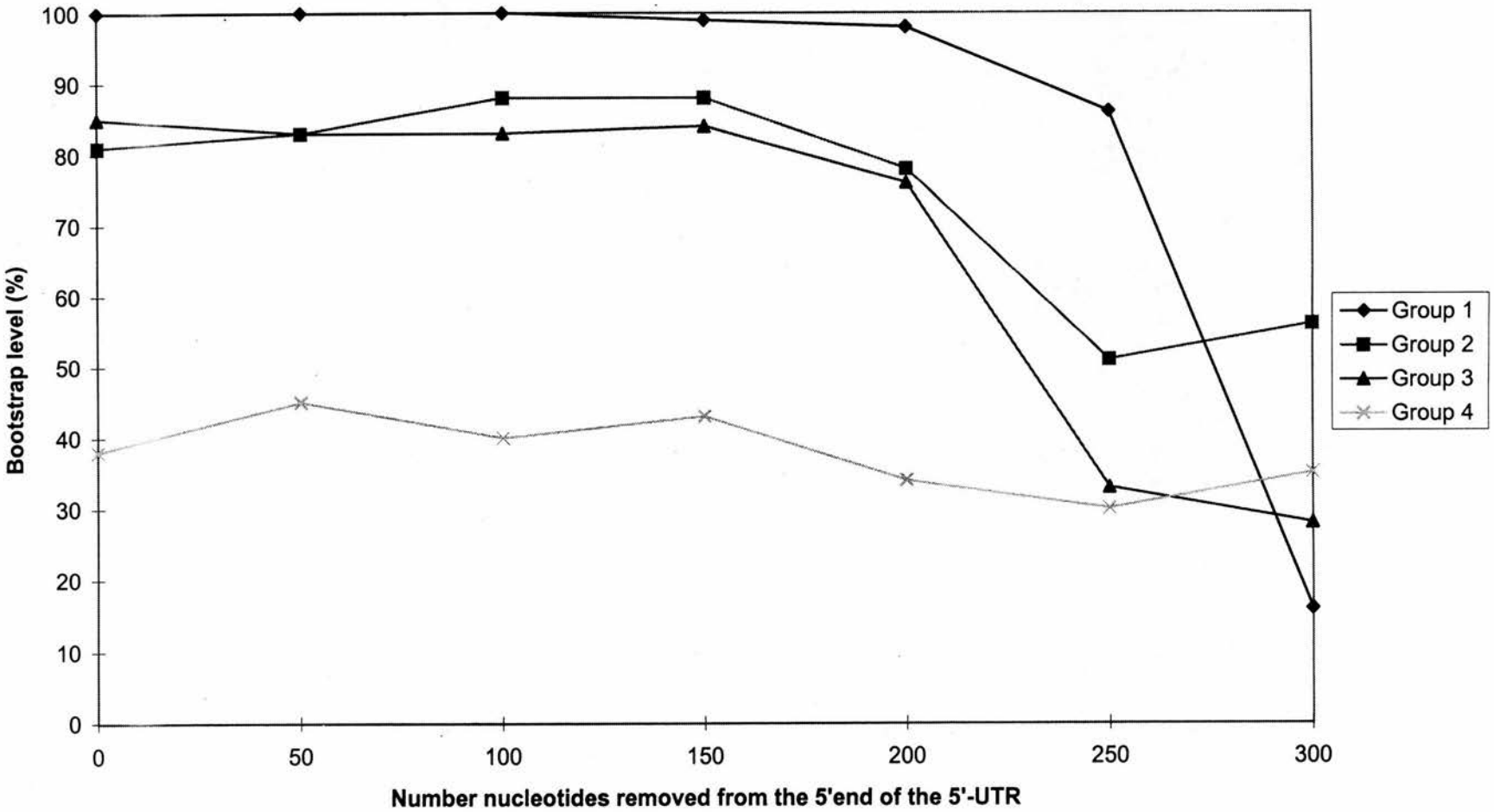
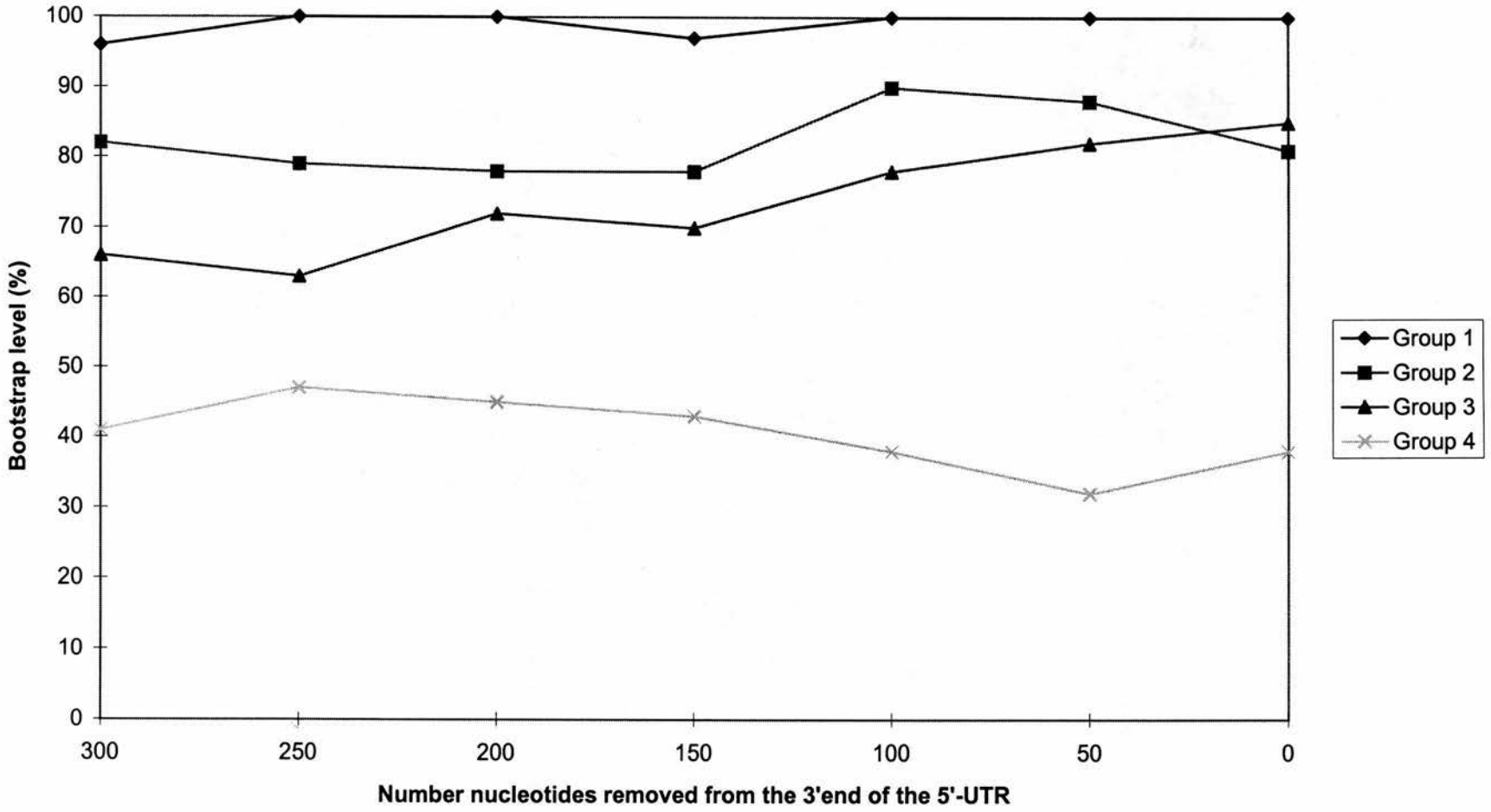

Bootstrap level (%)
100
90
80
70
60
50
40
30
20
10
0
0
50
100
150
200
250
300
Number nucleotides removed from the 5'end of the 5'-UTR
Group 1
Group 2
Group 3
Group 4

Figure 3.4 Bootstrap support for phylogenetic groups based on the analysis of progressively smaller fragments obtained by removal of nucleotides from the 3'-end of HGV/GBV-C 5'-UTR. The level of bootstrap support for each phylogenetic group was calculated using the MEGA package (p-distance) for 100 replicates and plotted against the number of nucleotides removed.


100
90
80
70
60
50
40
30
20
10
0
Bootstrap level (%)
300
250
200
150
100
50
0
Number nucleotides removed from the 3'end of the 5'-UTR
Group 1
Group 2
Group 3
Group 4

analysis was then extended using additional 5'-UTR fragments deleted at both 5'- and 3'-ends. Deletion of 50 nucleotides at the 3'-end and up to 200 nucleotides from the 5'-end did not affect the bootstrap support for groups 1-3 (over 70%); in contrast, the bootstrap level for groups 2 and 3 was low for a fragment obtained by removing 50 nucleotides from 3'-end and 250 nucleotides from 5'-end (Figure 3.5 A). At the same time, analysis of segments obtained by deletion of 50 nucleotides at 5'-end and up to 300 nucleotides at 3'-end revealed similar degree of bootstrap support for groups 1-4 as that based on analysis of complete 5'-UTR fragment (Figure 3.5 B). One exception was observed when analyzing a fragment obtained by removing 50 nucleotides from 5'-end and 200 nucleotides from 3'-end: the bootstrap support for group 3 was less than 70%. Shorter fragments obtained by removing 100 nucleotides from 5'-end and 250 nucleotides from 3'-end (Figure 3.5 C) or 150 nucleotides from 5'-end and 250 nucleotides from 3'-end or 200 nucleotides from 5'-end and 250 nucleotides from 3'-end (Figure 3.5 D) affected the level of bootstrap support, especially for group 3 (less than 70%). Surprisingly, deletion of 150 nucleotides from the 5'-end and 200 nucleotides from the 3'-end improved the bootstrap support for the phylogenetic groups 1-3, with levels over 90% for groups 1 and 2 and over 70% for group 3 (Figure 3.5 D). In conclusion, the phylogenetic groups 1, 2, 3 and 4 of HGV/GBV-C variants were consistently produced but only the first three were supported by bootstrap values over 70% when segments between positions -353 and -50 (Figure 3.5 A), -503 and -300 (Figure 3.5 B), -353 and -100 (Figure 3.5 C) or -403 and -200 (Figure 3.5 D) were considered.

Figure 3.5 Bootstrap support for the phylogenetic groups based on the analysis of fragments of the 5'-UTR of HGV/GBV-C complete genome sequences. Phylogenetic analysis was performed on progressively smaller 5'-UTR subfragments generated by removal of: A. 250 nucleotides at 5'-end and 50 nucleotides at 3'-end; B. 50 nucleotides at the 5'-end and 300 nucleotides at 3'-end; C. 250 nucleotides at 5'-end and 250 nucleotides at 3'-end; D. 200 nucleotides at 5'-end and 250 nucleotides at 3'-end. Nucleotide positions within the 5'-UTR are shown on the bottom of the figure.

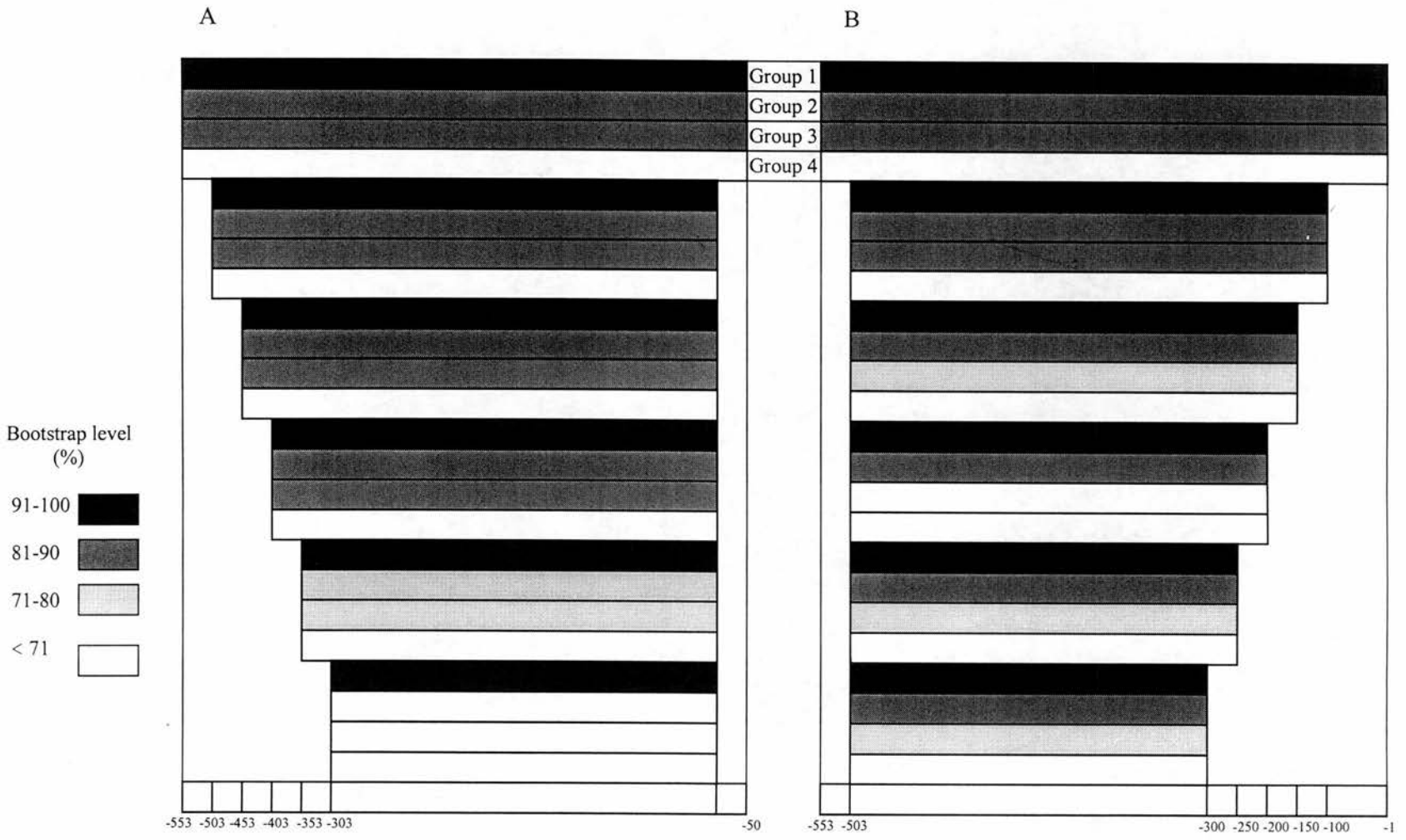
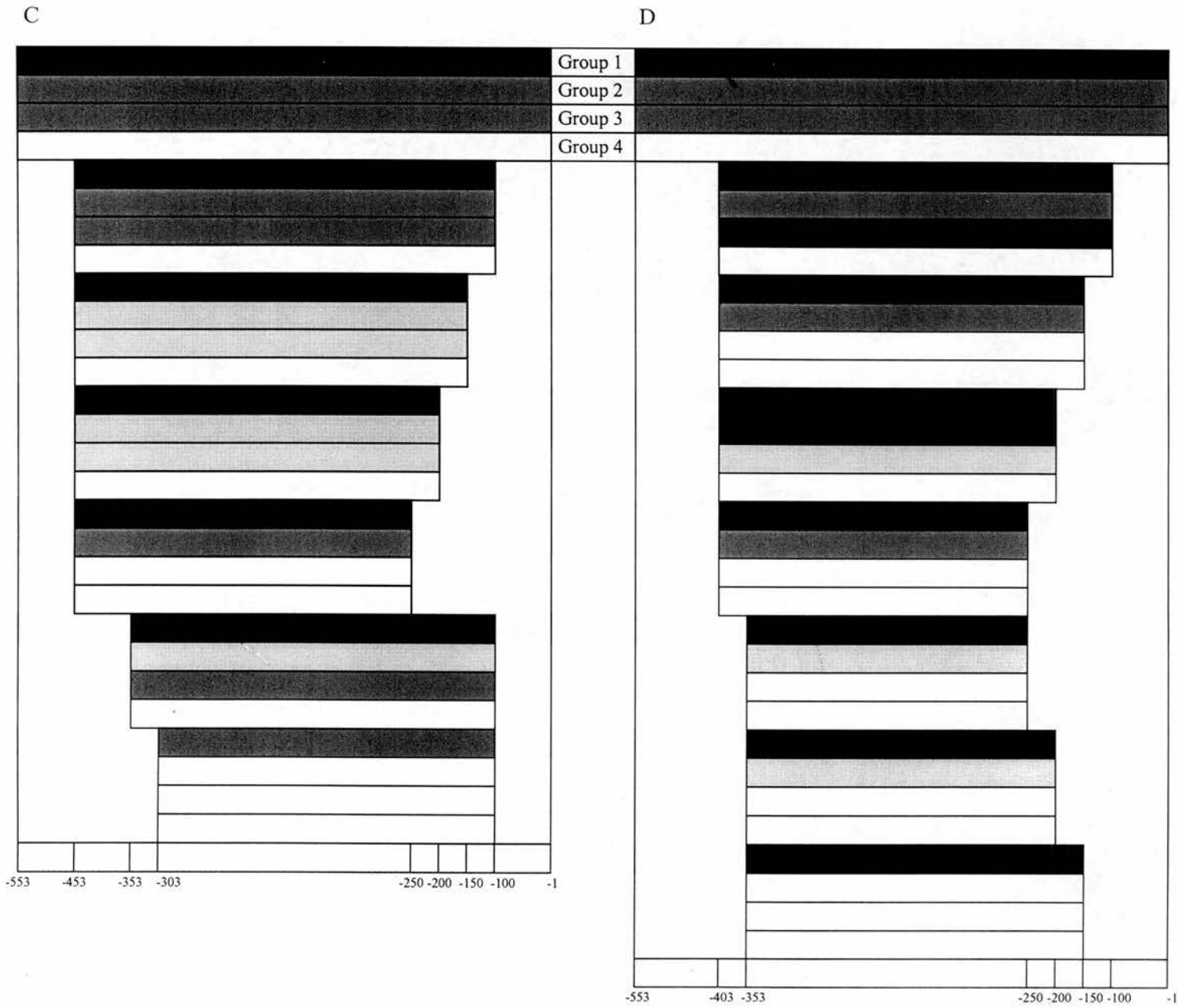

A
B
Group 1
Group 2
Group 3
Group 4
Bootstrap level (%)
91-100
81-90
71-80
< 71
-553 -503 -453 -403 -353 -303
-50
-553 -503
-300 -250 -200 -150 -100
-1

Figure 3.5


C
D
Group 1
Group 2
Group 3
Group 4
-553
-453
-353
-303
-250
-200
-150
-100
-1
-553
-403
-353
-250
-200
-150
-100
-1

Having defined the region of the 5'-UTR that optimally allows the discrimination of HGV/GBV-C phylogenetic groups (between positions -403 and -200 (Figure 3.5 D), this region was then used to genotype sequences obtained from samples from Zaïre (n=9), Pakistan (n=14) and Edinburgh (n=12) (Figure 3.6). HGV/GBV-C sequences from Zaïre clustered with the GBV-C prototype from West Africa in group 1, while HGV/GBV-C sequences from Pakistan clustered with sequences previously described as group 2 sequences. The sequences obtained from Edinburgh haemophiliac samples all clustered with group 2 isolates except for Ed_3 which clustered with group 1 and Ed_81 which clustered with group 3. These two haemophiliac patients (Ed_3 and Ed_81) had received commercial factor concentrates in contrast to the other haemophiliacs studied who had only received factor concentrates manufactured from blood donations collected in Scotland.

Bootstrap resampling of all the sequences for the region -366 to -235 again provided strong support for the existence of group 1 (97%), group 2 (85%) and group 3 (76%) but not for group 4 (Figure 3.7).

Figure 3.6 Comparison the HGV/GBV-C 5'-UTR sequences from Pakistan (Pak), Edinburgh (Ed) and Zair (Zai) with the 5'-UTR sequences of the HGV/GBV-C complete genome isolates. The alignment was generated manually. Sequence identity with GT110 is indicated by dots and the missing information by spaces. The group number is indicated at the right.

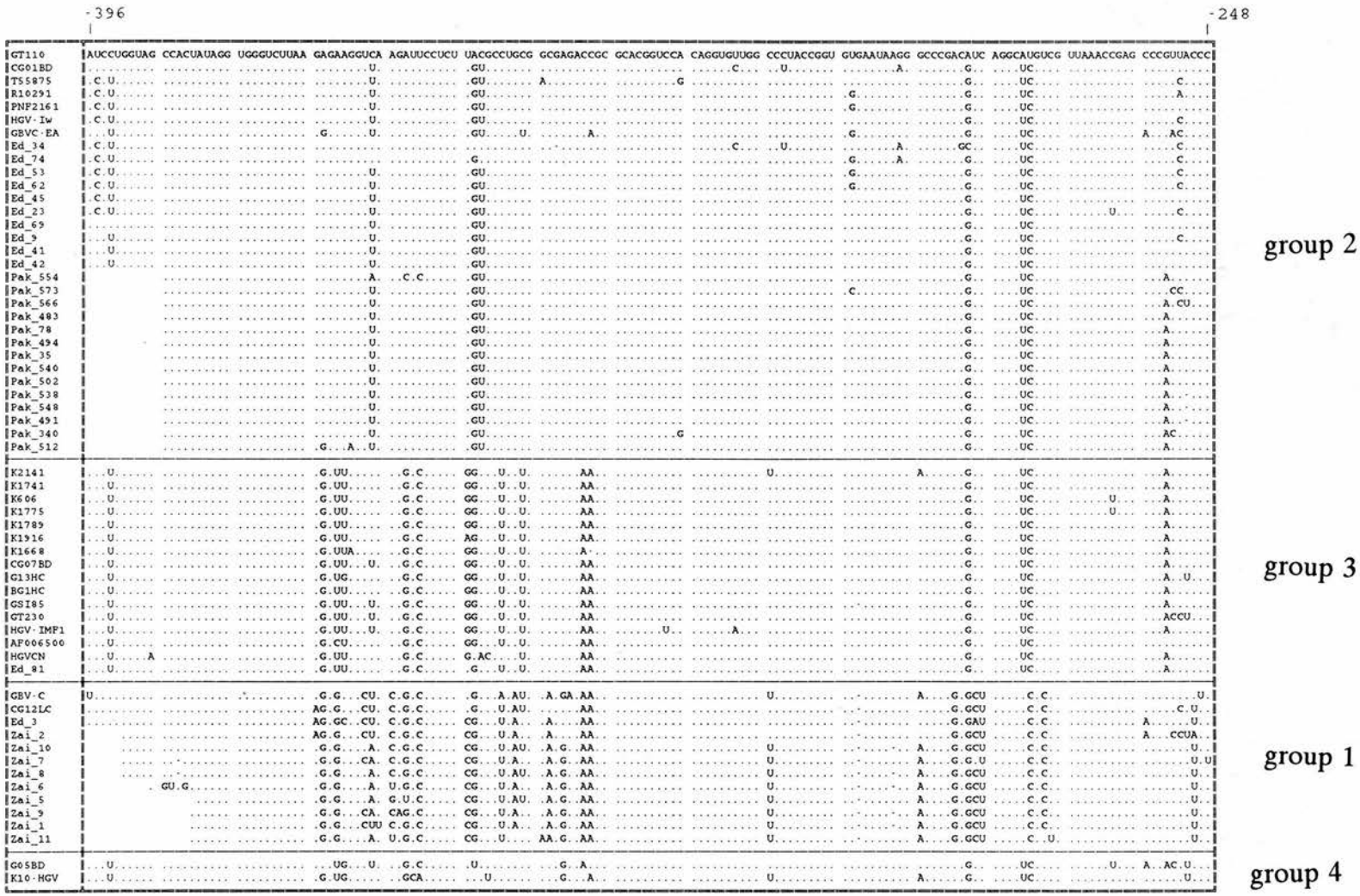

-396
-248
GT110 AUCCUGGUAG CCACUAUAGG UGGGUCUUAA GAGAAGGUCA AGAUUCCUCU UACGCCUGCG GCGAGACCGC GCACGGUCCA CAGGUGUUGG CCCUACCGGU GUGAAUAAGG GCCCGACAUC AGGCAUGUCG UUAAACCGAG CCCGUUACCC
CG01BD
T55875
R10291
PNF2161
HGV-Iw
GBVC-EA
Ed_34
Ed_74
Ed_53
Ed_62
Ed_45
Ed_23
Ed_69
Ed_9
Ed_41
Ed_42
Pak_554
Pak_573
Pak_566
Pak_483
Pak_78
Pak_494
Pak_35
Pak_540
Pak_502
Pak_538
Pak_548
Pak_491
Pak_340
Pak_512
group 2
K2141
K1741
K606
K1775
K1789
K1916
K1668
CG07BD
G13HC
BG1HC
GSI85
GT230
HGV-IMF1
AF006500
HGVCN
Ed_81
group 3
GBV-C
CG12LC
Ed_3
Zai_2
Zai_10
Zai_7
Zai_8
Zai_6
Zai_5
Zai_9
Zai_1
Zai_11
group 1
G05BD
K10-HGV
group 4

Figure 3.6 continued

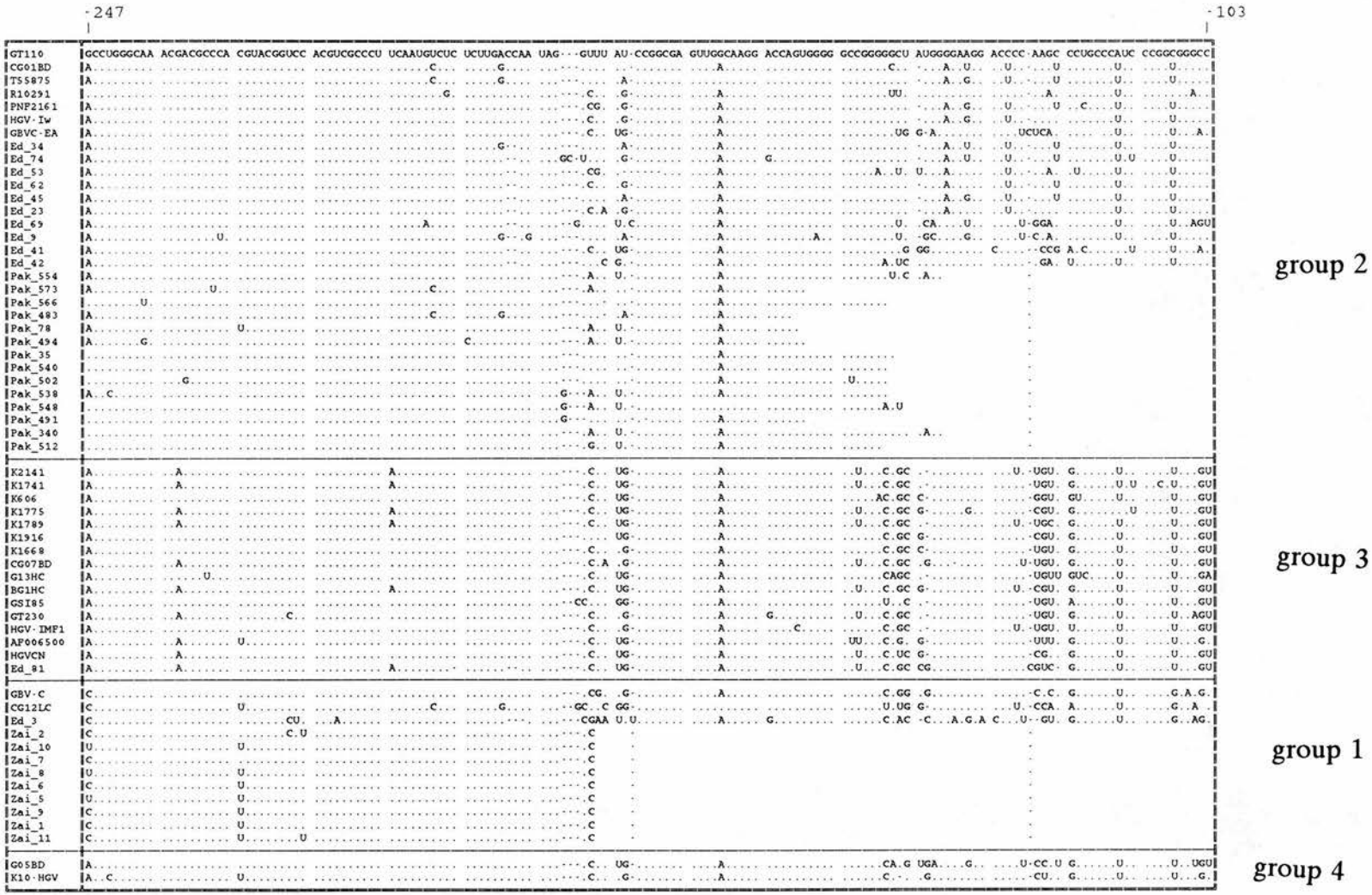


Figure 3.6 continued

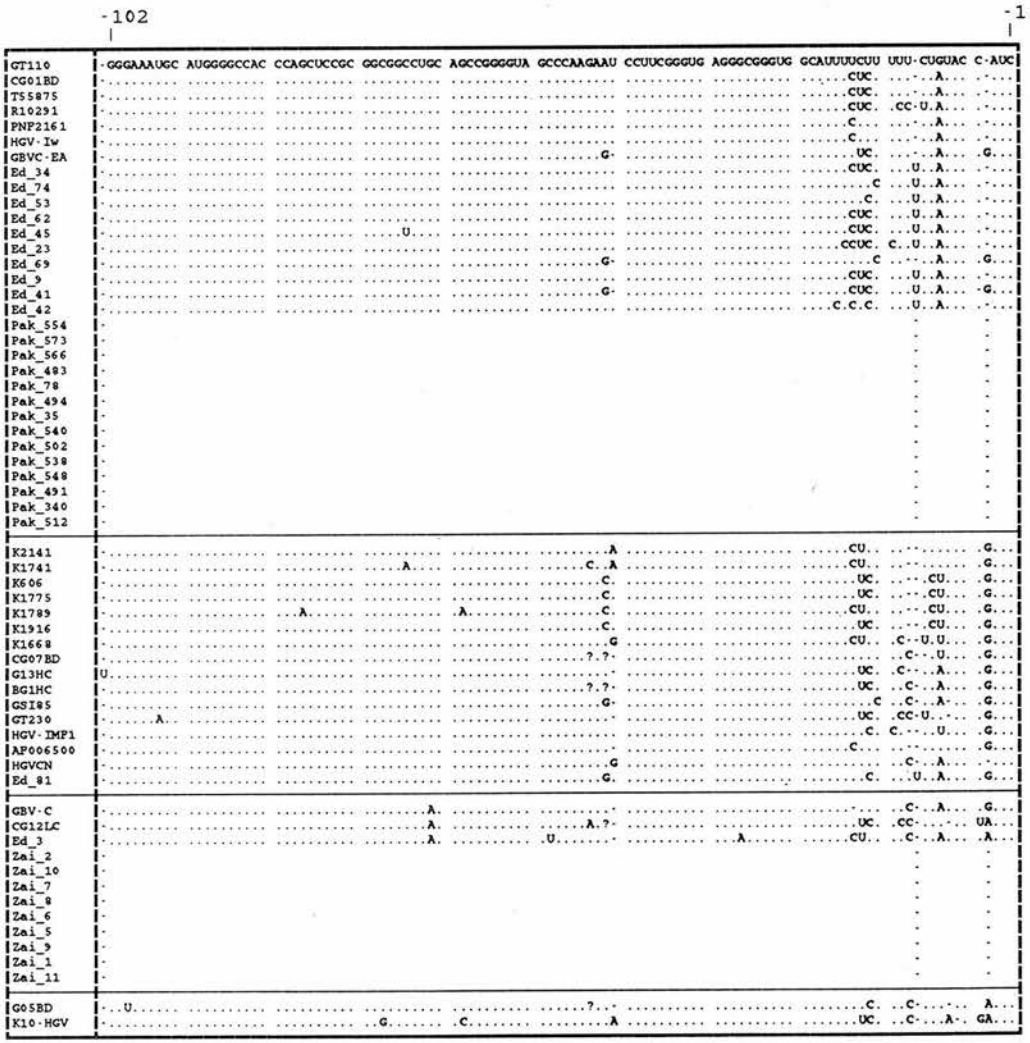
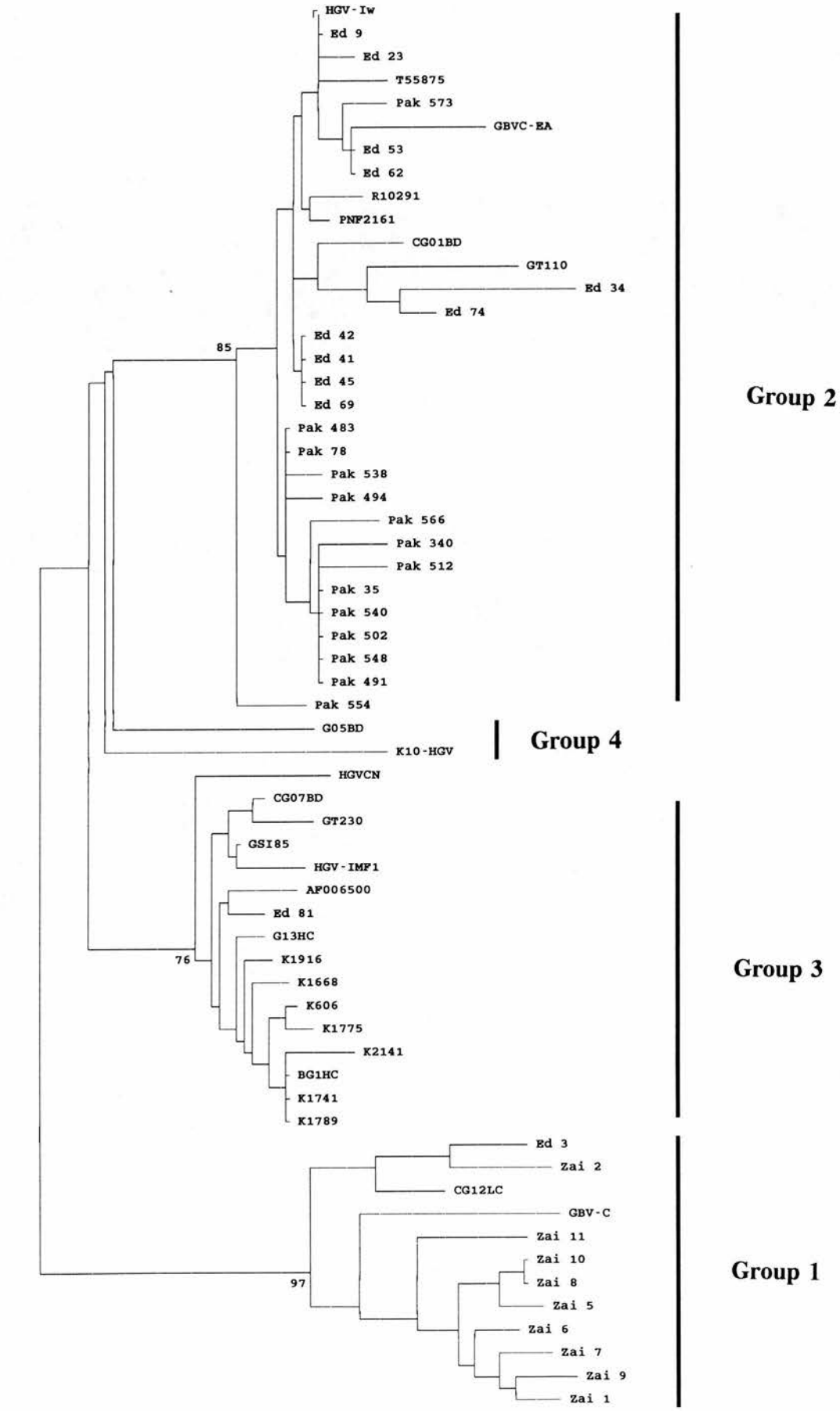


Figure 3.7 HGV/GBV-C phylogenetic groupings generated by comparison of 5'-UTR fragments between positions -366 and -235. A consensus phylogenetic tree was constructed from p-distance of the HGV/GBV-C 5'-UTR sequences by neighbor-joining method. Bootstrap values greater than 70% are indicated. The group number corresponding to the phylogenetic analysis of the HGV/GBV-C complete genome sequences is shown on the right.

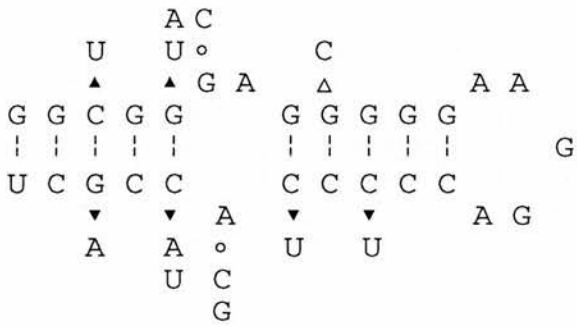

HGV-Iw
Ed 9
Ed 23
T55875
Pak 573
GBVC-EA
Ed 53
Ed 62
R10291
PNF2161
CG01BD
GT110
Ed 34
Ed 74
85
Ed 42
Ed 41
Ed 45
Ed 69
Pak 483
Pak 78
Pak 538
Pak 494
Pak 566
Pak 340
Pak 512
Pak 35
Pak 540
Pak 502
Pak 548
Pak 491
Pak 554
G05BD
K10-HGV
HGVCN
CG07BD
GT230
GSI85
HGV-IMF1
AF006500
Ed 81
G13HC
76
K1916
K1668
K606
K1775
K2141
BG1HC
K1741
K1789
Ed 3
Zai 2
CG12LC
GBV-C
Zai 11
Zai 10
Zai 8
97
Zai 5
Zai 6
Zai 7
Zai 9
Zai 1
Group 2
Group 4
Group 3
Group 1
Scale: is approximately equal to the distance of 0.02

The sequence alignment of the HGV/GBV-C 5'-UTR sequences revealed the presence of sequence polymorphisms throughout the region, some of which were associated with particular phylogenetic groupings (Figures 3.2, 3.6). For example, clusters of group-specific polymorphisms are located at positions -104, -103, between positions -366 to -329, -287 and -235. An additional cluster of group-specific 5'-UTR polymorphisms occurs between positions -490 and -459 but relatively little sequence information is available for this region and some groups are not represented (Figure 3.2). Other substitutions present within 5'-UTR are not correlated with phylogenetic groupings. For example, substitutions located between positions -17 and -15 (within the polypyrimidine stretch) and at positions -115 and -167 occur in all phylogenetic groups (Figure 3.2).

The majority of substitutions observed were consistent with proposed secondary structures for HGV/GBV-C 5'-UTR (Simons *et al.* 1996b; Pickering *et al.* 1997b) since they either were present in unpaired regions (e.g., between positions -490 and -485, -341 and -329) or did not disrupt the base-pairings. For example, a stem-loop structure between positions -147 and -120 corresponding to region IVb in the proposed structure of Simons et al. (Simons *et al.* 1996), could be predicted for all the HGV/GBV-C sequences from this study, despite the presence of many substitutions within this region (Figure 3.8). Although some of these substitutions are group-specific, most of them occur in all phylogenetic groups.

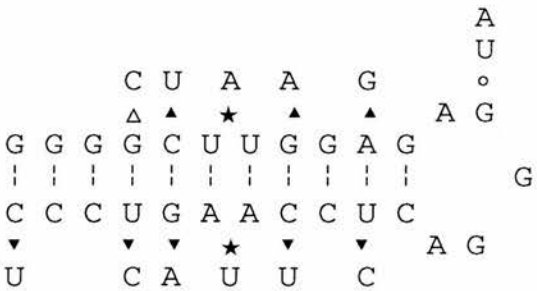
Figure 3.8 Predicted stem-loop structures (positions -147 to -120) for different HGV/GBV-C groups. Substitutions do not affect (▲) or disrupt (△) base pairings,or occur in unpaired regions (○); (★) nucleotide insertion. The range of free energy (kcal) amongst variants of each group is indicated in brackets.

GROUP 1



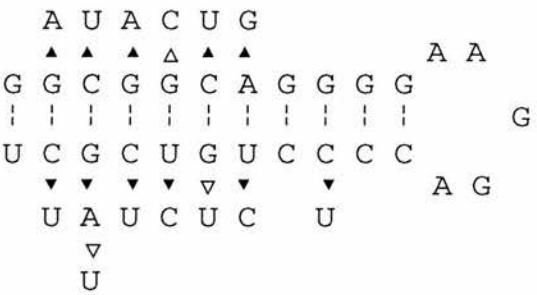
(-16.55 to -17.30)

GROUP 2



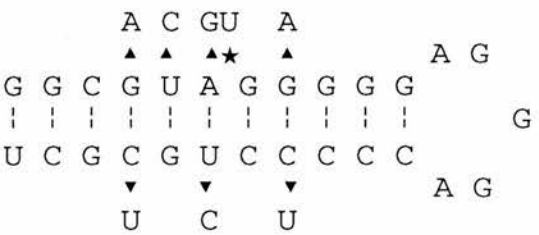
(-9.17 to -18.33)

GROUP 3



(-8.65 to -18.68)

GROUP 4



(-12.96 to -19.19)

3.4 DISCUSSION

COMPLETE GENOME SEQUENCE ANALYSIS

The phylogenetic analysis of twenty-seven full-length or nearly full-length HGV/GBV-C nucleotide sequences has revealed the existence of four phylogenetic groups with bootstrap values of 100% for each of them (Figure 3.1). This result is consistent with but extends the previously published studies based on comparative analysis of complete genome HGV/GBV-C sequences which suggested the existence of only three groups (Nakao *et al.* 1997; Okamoto *et al.* 1997; Katayama *et al.* 1998). The isolates that grouped together by analysis of the full-length sequences also clustered together by phylogenetic analysis of the entire 5'-UTR or of various shorter fragments of the 5'-UTR (Figures 3.3, 3.4 and 3.5).

One anomaly of this analysis was the Chinese isolate HGVC 964 that segregated with group 2 isolates when the 5'-UTR sequence was analysed but with group 3 isolates upon analysis of the complete genome sequence (Figure 3.1). However, the genome sequence of this variant is 78 nucleotides shorter than the GBV-C prototype sequence (Muerhoff *et al.* 1997), with deletions/insertions present throughout the coding regions and the 3'-UTR is unrelated to any sequence from database. Because of these discrepancies, and the lack of supporting published information, HGVC 964 was excluded from the phylogenetic analysis of the 5'-UTR fragments.

Analysis of the 5'-UTR or of subfragments failed to provide bootstrap support (less than 70% of replicates) for group 4 (G05BD and K10-HGV isolates). G05BD isolate was previously considered as a potentially new genotype following the phylogenetic analysis of approximately 2kb-sequence that covered entire E1 and most of 5'-UTR and E2 regions of HGV/GBV-C genome (Takahashi *et al.* 1997). When the phylogenetic analysis was extended to the complete polyprotein coding region, G05BD clustered with type 3 isolates only for analysis of the NS5B region (Takahashi *et al.* 1997). Although the existence of group 4 isolates looks uncertain at present, the identification of a novel group of HGV/GBV-C variants based on analysis of 5'-UTR sequences from Thailand isolates support this observation (Katayama *et al.* 1997). Comparison of these nucleotide sequences with G05BD and K10-HGV sequences indicated a high degree of identity (data not shown) but the overlapping fragments between them were too short to analyse further their phylogenetic relationships. This issue would be better investigated when the full-length or longer sequences of the Thai isolates are available.

In conclusion, the phylogenetic groups of HGV/GBV-C variants identified upon analysis of complete genome sequences can be distinguished by analysis of the entire 5'-UTR or of various subfragments (Figure 3.5 A, B, C, D). The optimal region for identifying the groups appeared to be the region between positions -403 and -200.

Two of the three major HGV/GBV-C phylogenetic groupings have been divided into subgroups (1a, 1b, 2a, 2b) based on the comparison of the 5'-UTR

sequences (Muerhoff *et al.* 1996b). Evolutionary distances between types 1, 2 or 3 were approximately 0.1000, while distances between variants of the same type were approximately 0.034 to 0.056. At the same time, in an attempt to determine the phylogenetic groupings of 16 HGV/GBV-C isolates obtained from multitransfused patients a novel subgroup, named 2c, was proposed based on phylogenetic analysis of the entire 5'-UTR (LopezAlcorocho *et al.* 1999). Comparison of evolutionary distances indicated that the HGV/GBV-C sequences isolated from most patients belonged to subgroup 2a while the phylogenetic trees constructed by analysis of the entire 5'-UTR or a 320bp-fragment of 5'-UTR showed that 5 of the isolates segregated into a separate branch, close to those of subgroup 2a and 2b sequences. However, since the bootstrap support for this segregation is not very strong (70% of replicates), these ambiguous results might be the result of errors in the sequence alignment. In another published extended phylogenetic analysis of 5'-UTR and coding region sequences of HGV/GBV-C variants, the bootstrap support for segregation of group 1 isolates into subgroups was not very strong while segregation of group 2 sequences was difficult to obtain unless a significant fragment of the 5'-UTR was analysed (Muerhoff *et al.* 1997). Based on these findings and because the complete genome sequence analysis carried out in this chapter included only one isolate (GBVC-EA) previously defined as subgroup 2b, only the major HGV/GBV-C phylogenetic groupings were considered for discussion.

In contrast to analysis of subgenomic coding regions, phylogenetic analysis of fragments of 5'-UTR can be used to define the phylogenetic groupings of various HGV/GBV-C variants which are well correlated with their geographical origin, although there is no perfect correlation between the country and phylogenetic group. This conclusion was supported by analysis of 61 different sequences carried out in this chapter (Figure 3.7) where the following groups could be distinguished: Group 1 contains African isolates, Group 2 consists mainly of isolates from Europe, North America and Indian subcontinent as well as GBVC-EA from East Africa and HGV-Iw from Japan, and Group 3 is predominant in Asia (Japan). Two sequences (G05BD and K10-HGV) may represent a new group (Group 4) since analysis of complete genome sequences indicates that it is a distinct group from Groups 1, 2, and 3 (Figure 3.1).

At the same time, the geographical distribution of HGV/GBV-C phylogenetic groups and their diversity and prevalence in different human populations might offer information on HGV/GBV-C origin. Recently, this issue was investigated by phylogenetic analysis of the NS3 and NS5A regions from 94 HGV/GBV-C isolates collected from Central Africa and other areas (e.g., Australia, Japan, Uzbekistan) (Tanaka *et al.* 1998). The prevalence of HGV/GBV-C was high in Central African region where most of the isolates belonged to Group 1 (95.65%); 15 of Group 1 isolates had a 12-amino acid indel (i.e. insertion or deletion) in the NS5A region. Phylogenetic analysis of the

NS5A region in which GBV-A was used as an outgroup, revealed that the 15 isolates (Indel type) formed a separate outer group suggesting that they had diverged earlier than the Non-Indel sequences from a common ancestor. HGV/GBV-C might be of African origin and its transmission to new communities may perhaps taken place along with human migration. However, this hypothesis remains to be further analysed.

VARIATION OF HGV/GBV-C 5'-UTR

In general, a high degree of conservation was identified throughout the entire genome of HGV/GBV-C (Erker *et al.* 1996; Mukaide *et al.* 1997; Okamoto *et al.* 1997; Wang *et al.* 1997). The 5'-UTR consists of both well conserved and variable regions (Figures 3.2, 3.6). The alignment of 5'-UTR sequences indicated the presence of nucleotide polymorphisms (e.g., between positions -490 and -235) that were associated with particular phylogenetic groups. As previously suggested (Muerhoff *et al.* 1996b), they could be used to segregate new HGV/GBV-C sequences into groups. This observation is supported by the phylogenetic tree of different HGV/GBV-C isolates constructed by comparison of 5'-UTR fragments located between positions -366 and -235 (Figure 3.7).

In case of HCV, an association between particular polymorphisms in the 5'-UTR and virus genotype has led to the development of HCV genotyping methods by line probe assay (LIPA) (Stuyver *et al.* 1993) or by restriction

fragment length polymorphism analysis (RFLP) (McOmish *et al.* 1993; McOmish *et al.* 1994; Davidson *et al.* 1995). Similarly, the presence of polymorphisms in the HGV/GBV-C 5'-UTR provided also evidence for the identification of group-specific restriction sites for 2 restriction enzymes (ScrF I and BsmF I (Mukaide *et al.* 1997). For example, the sequence recognized by BsmF I (GUCCC) for group 1 and group 3 HGV/GBV-C isolates was found located between positions -354 and -350.

At the same time, many of the polymorphisms are covariant and consistent with the predicted secondary structure for the 5'-UTR which has a potential functional importance in different viral events such as the initiation of translation (Simons *et al.* 1996). However, the validity of the model remains to be confirmed by ribonuclease sensitivity assays.

The extensive phylogenetic analysis carried out in this chapter on HGV/GBV-C complete genome sequences and 5'-UTR sequences revealed significant aspects of HGV/GBV-C genetic heterogeneity, suggesting some guidelines for assigning the phylogenetic groups.

Classification of HGV/GBV-C isolates in groups is important not only for evaluating the geographical distribution and diversity of this virus or for providing basis for the development of reliable detection methods but also for understanding its evolution within the human population.

CHAPTER 4

4. BUOYANT DENSITY AND SEDIMENTATION BEHAVIOUR OF HGV/GBV-C

4.1 INTRODUCTION

The genome organization of HGV/GBV-C is relatively similar to that of HCV (Leary *et al.* 1996; Linnen *et al.* 1996) but it is not currently known if these similarities extend to the virus structure and morphological properties.

In previous studies, the HCV particle has been visualized by immuno-electron microscopy and estimated to be 55 to 65 nm in diameter with spike-like projections 6 nm long (Kaito *et al.* 1994). Similarly, filtration studies have indicated that virus particles have a diameter of 30-60 nm (He *et al.* 1987; Yuasa *et al.* 1991). These values are within the range of estimated diameters of flaviviruses (35 and 50 nm) (Westaway *et al.* 1985). Lipid solvents such as chloroform reduce the infectivity of HCV (Bradley *et al.* 1983; Feinstone *et al.* 1983) and disrupt the virus envelope (Hijikata *et al.* 1993) indicating that the virus has an envelope containing lipids.

Early studies reported the buoyant density of HCV virus particles in sucrose at 1.08 g/ml (Miyamoto *et al.* 1992) and 1.09-1.11 g/ml (Bradley *et al.* 1991), similar to the density of the pestiviruses bovine viral diarrhoea virus (1.09-1.15 g/ml) and hog cholera virus (1.12-1.16 g/ml) but lower than those of most flaviviruses (1.19-1.20 g/ml) (Westaway *et al.* 1985). The HCV present in fractions with low density (1.03-1.12 g/ml) is infectious (Bradley *et al.* 1991; Hijikata *et al.* 1993) and is precipitated by goat anti-human β lipoprotein

antibodies, indicating an association with plasma low-density lipoproteins (LDL) (Bradley *et al.* 1991; Thomssen *et al.* 1993). However, subsequent studies have demonstrated the presence in serum of a higher density fraction (1.17-1.20 g/ml) in which virus is bound to anti-viral antibodies in immune complexes that can be precipitated by rabbit anti-human immunoglobulins directed against subtypes IgA, IgG and IgM (Hijikata *et al.* 1993; Kanto *et al.* 1995) or by anti-IgG (Thomssen *et al.* 1993). An intermediate fraction with density range from 1.10 to 1.16g/ml present in some human sera contains a mixture of viral lipoproteins, viral aggregates and incomplete virus particles (Thomssen *et al.* 1993). HCV RNA is also present in fractions with densities exceeding 1.21 g/ml in the sera of chronic hepatitis patients (Thomssen *et al.* 1992) and appear to consist of nucleocapsids since they can be immunoprecipitated with monoclonal anti-core HCV antibody (Kanto *et al.* 1995). A fraction with a similar density containing HCV nucleocapsids (1.25 g/ml) was isolated after detergent treatment from plasma of blood donors (Miyamoto *et al.* 1992). Studies of HCV density distribution in patients with various stages of chronic hepatitis suggested that the dominant HCV population shifted from low-density to high-density particles with the progression of liver disease or increased inflammation (Kanto *et al.* 1995).

At the same time, antibodies to the HVR-1 region may be responsible for association of HCV virions with immunoglobulins since they could be detected in infected plasma by ELISA using overlapping oligopeptides deduced from HVR of HCV (Kojima *et al.* 1994). Within E2 sequence encoding for one of the

major structural glycoproteins of HCV, there are regions of extreme hypervariability (HVR); one of these-HVR1-which represents the amino terminal 25-30 amino acids of the E2 protein appeared to be structurally flexible and antigenically variable during a HCV infection follow-up, due to a significant number of missense mutations (Taniguchi *et al.* 1993). These successive antigenic changes of HVR1 may provide the virus a way to escape from humoral immune response of the host.

Based on the sequence comparisons with HCV, homologues to the E1 and E2 envelope proteins have also been identified in HGV/GBV-C but these do not contain HVR regions (Okamoto *et al.* 1997). Although these proteins may be glycosylated (Leary *et al.* 1996b), only three N-linked glycosylation sites have been identified in E2 of HGV/GBV-C while there are between eight and eleven in E2 protein of HCV (Leary *et al.* 1996; Okamoto *et al.* 1997).

Another important difference between the polyproteins of HGV/GBV-C and HCV is the absence of a putative HGV/GBV-C core protein, which is usually encoded at the 5'end of the genome. The HCV core protein is a highly basic protein of 191 amino acids located at the amino-terminus of the polyprotein that is well conserved amongst HCV genotypes 1-6 (Cha *et al.* 1992; Bukh *et al.* 1994). Its biosynthesis and biochemical properties have been investigated and specific functions mapped to discrete portions of the molecule (Santolini *et al.* 1994; Matsumoto *et al.* 1996). For example, the core protein released from the HCV polyprotein precursor is a cytoplasmic protein associated with the endoplasmic reticulum (ER) membranes and possesses RNA binding

activity. The ribosome- and RNA-binding domain were mapped to the NH_2-terminus of the core protein between amino acids 1 and 75 while the C-terminal hydrophobic region was found to be responsible for the membrane association of the HCV core protein (Santolini *et al.* 1994). At the same time, the homotypic interaction and multimerization of the HCV core protein *in vitro* and *in vivo* was demonstrated (Matsumoto *et al.* 1996). The NH_2-terminal hydrophilic domain (amino acids 1-115) is mainly responsible for these events. In contrast, although many full-length HGV/GBV-C genomes have been published, no consistent viral nucleocapsid or core protein sequence has been revealed so far (Erker *et al.* 1996; Leary *et al.* 1996; Linnen *et al.* 1996; Okamoto *et al.* 1997; Katayama *et al.* 1998). However, a putative E1 signal peptidase recognition sequence and an Asn-Cys-Cys motif within E1 protein that is highly conserved in HCV, GBV-A, GBV-B and HGV/GBV-C have been identified (Leary *et al.* 1996b). In addition, *in vitro* translation studies of transcripts containing the 5'-end of the GBV-C genome indicate that the site of translation initiation is located immediately upstream of a putative E1 signal sequence. This implies that a core-like protein does not exist at the amino-terminus of the polyprotein (Simons *et al.* 1996).

Together, these discrepancies in the genome organisation of HGV/GBV-C suggest that the morphology and physicochemical properties of HGV/GBV-C particles may be different from those of HCV. This chapter describes an investigation of the buoyant density and the sedimentation behaviour of HGV/GBV-C particles from human plasma in comparison with HCV.

Three HCV positive plasma samples (A, B, C) and three HGV/GBV-C positive plasma samples (F, G, H) were used in this study. Samples A and B were anti-HCV positive while sample C was seronegative on screening by Abbott second generation enzyme immunoassay (EIA) and confirmed by RIBA-3 assay. Another plasma sample (N), that was both HCV and HGV/GBV-C negative, was used as negative control (Table 4.1). The samples (500μl) were fractionated on a preformed density gradient ranging from 10% to 60% (w/w) sucrose in TEN buffer at 27,800 rpm for 44 h at 10^0C in a Beckman SW50.1 ultracentrifuge rotor (section 2.1).

The presence of viral RNA in each fraction was detected by RT-PCR using nested primers specific for the 5'-UTR of HCV and HGV/GBV-C, respectively (section 2.2.4). The quantification of viral RNA was performed using the limiting dilution method in which cDNA is titrated to the endpoint for detection by PCR amplification (section 2.2.7); the number of copies of RNA per ml of plasma was calculated assuming a 5% efficiency of the reverse transcription reaction as previously established for HIV (Zhang *et al.* 1991). The sensitivity of detection was approximately 10^3 copies of RNA/ml.

HCV RNA was widely distributed over sucrose density fractions with densities ranging from 1.04 to 1.21 g/ml (Table 4.2). Quantification of viral RNA by limiting dilution revealed high levels of HCV RNA at densities of 1.04, 1.06 and 1.16 g/ml for sample A (Figure 4.1), 1.05, 1.12 and 1.19 g/ml for sample B

Table 4.1 Plasma samples used in the density analysis of HCV and HGV/GBV-C in sucrose gradients.

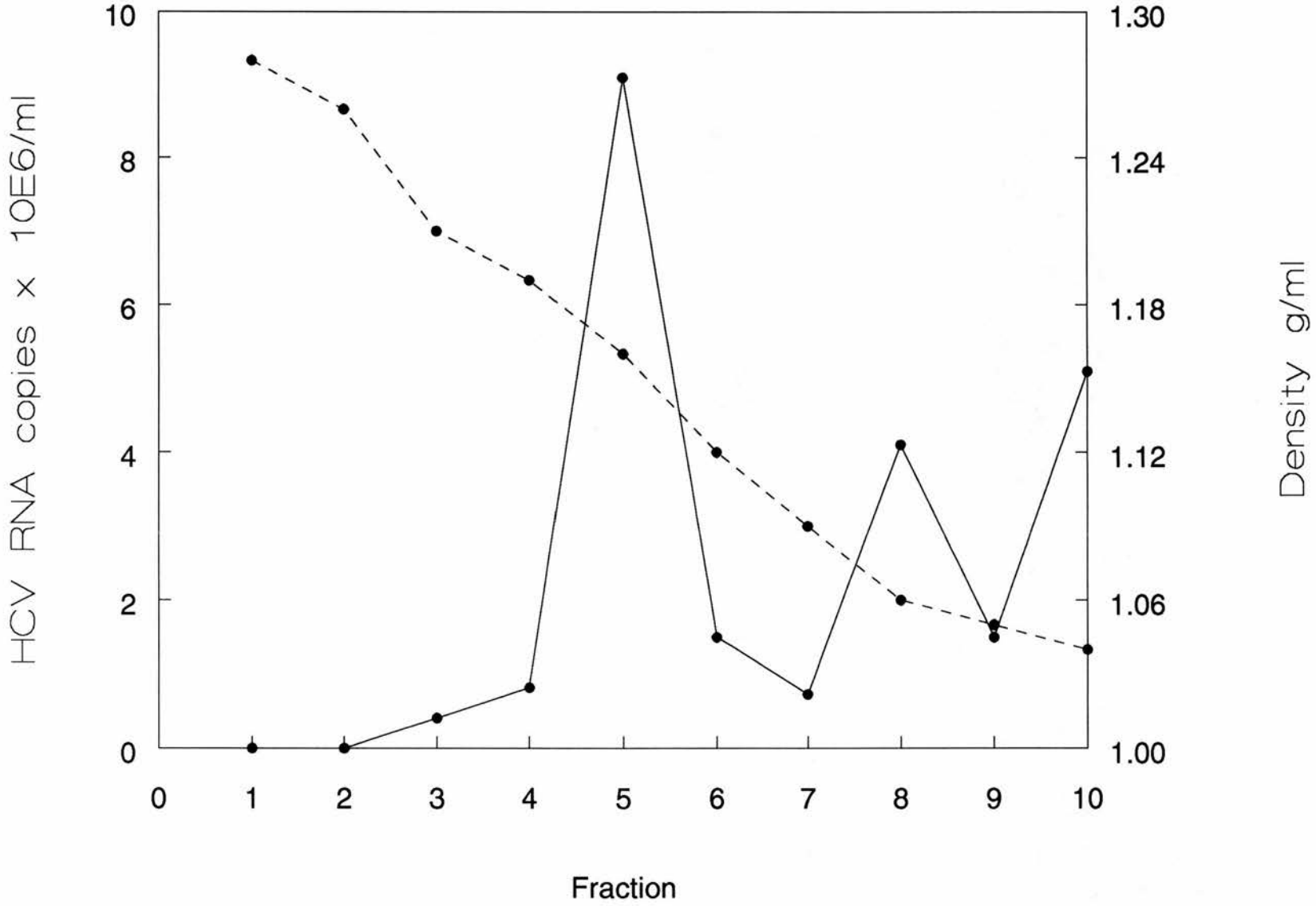
Sample	Patient	HCV RNA (genotype)	HGV/GBV-C RNA	anti-HCV Abs (ELISA+RIBA-3)
A	haemophiliac	+ (3)	-	+
B	blood donor	+ (1)	-	+
C	blood donor	+ (3)	-	-
F	haemophiliac	-	+	-
G	haemophiliac	-	+	-
H	haemophiliac	+ (1)	+	+
N	blood donor	-	-	-

Table 4.2 Presence of HCV and HGV/GBV-C RNA in sucrose fractions.

Fraction density	Sample / titre (no.copies RNA x 10^6/ml plasma)						
	HCV			HGV/GBV-C			Negative
g/ml	A	B	C	F	G	H	N
1.28	-			-			-
1.26	-	-		-		-	-
1.23		-	-		-	-	-
1.21	0.41	0.16	-	-	-		
1.20		0.33					
1.19	0.82	2.2	0.16	-		-	-
1.17					-	0.16	-
1.16	9.1			-			
1.15		-	1.31		-	0.25	
1.14							-
1.13			-		0.08		
1.12	1.5	1.76		0.51	0.44	0.88	-
1.11			0.33				
1.10					0.88		
1.09	0.73		0.82	0.91	1.76	0.66	-
1.08		0.66	8.8				
1.07		-				0.66	-
1.06	4.1		-	-		-	-
1.05	1.5	1.76		0.16	0.08	-	
1.04	5.1		1.31	0.2			
1.03					-		

For the PCR-negative samples , the titre $< 10^3$ copies/ml.

Figure 4.1 Sedimentation profile of sample A in sucrose gradient. Levels of viral RNA (solid line) were detected by limiting dilution RT-PCR assay. Dotted line shows variation of the sucrose density.


HCV RNA copies x 10E6/ml
Density g/ml
Fraction
10
8
6
4
2
0
1.30
1.24
1.18
1.12
1.06
1.00
0
1
2
3
4
5
6
7
8
9
10

(Figure 4.2) and 1.04, 1.08 and 1.15 g/ml for sample C (Figure 4.3). The density at which the peak of HCV RNA concentration was found was 1.16 g/ml for sample A, 1.19 g/ml for sample B and 1.08 g/ml for sample C. These results are similar to those reported in previous density centrifugation studies of serum samples from HCV-infected individuals in which the RNA was detected in both light (1.03-1.08 g/ml) and heavy (1.17-1.21 g/ml) density fractions (Miyamoto *et al.* 1992; Hijikata *et al.* 1993; Thomssen *et al.* 1993).

In contrast, the distribution of HGV/GBV-C RNA in the sucrose fractions of each HGV/GBV-C positive sample was more restricted, with only single peaks of virus RNA concentration (Table 4.2). The buoyant density at which the majority of HGV/GBV-C RNA was detected was between 1.07 and 1.12 g/ml according to the sample; the peak values of virus RNA were found at 1.09 g/ml for samples F (Figure 4.4) and G (Figure 4.5) and 1.12 g/ml for sample H (Figure 4.6). No HGV/GBV-C RNA was detected in fractions with densities higher than 1.17 g/ml. Two of the samples (B and F) were subjected to repeated sucrose density fractionation with similar results (data not shown) while a sample from a HCV and HGV/GBV-C negative individual was negative for all gradient densities for both viruses.

In order to obtain a more extensive characterization of HGV/GBV-C particles, an attempt to visualize the virus particles by electron microscopy was made. For this purpose, two sucrose fractions, each of sample A (at 1.04 and 1.16 g/ml) and of sample F (at 1.09 and 1.12 g/ml) were studied. 200 μl of each fraction were diluted in PBS and centrifuged at 15,000 rpm for 2.5 h at 4^0C. The

Figure 4.2 Sedimentation profile of sample B in sucrose gradient.

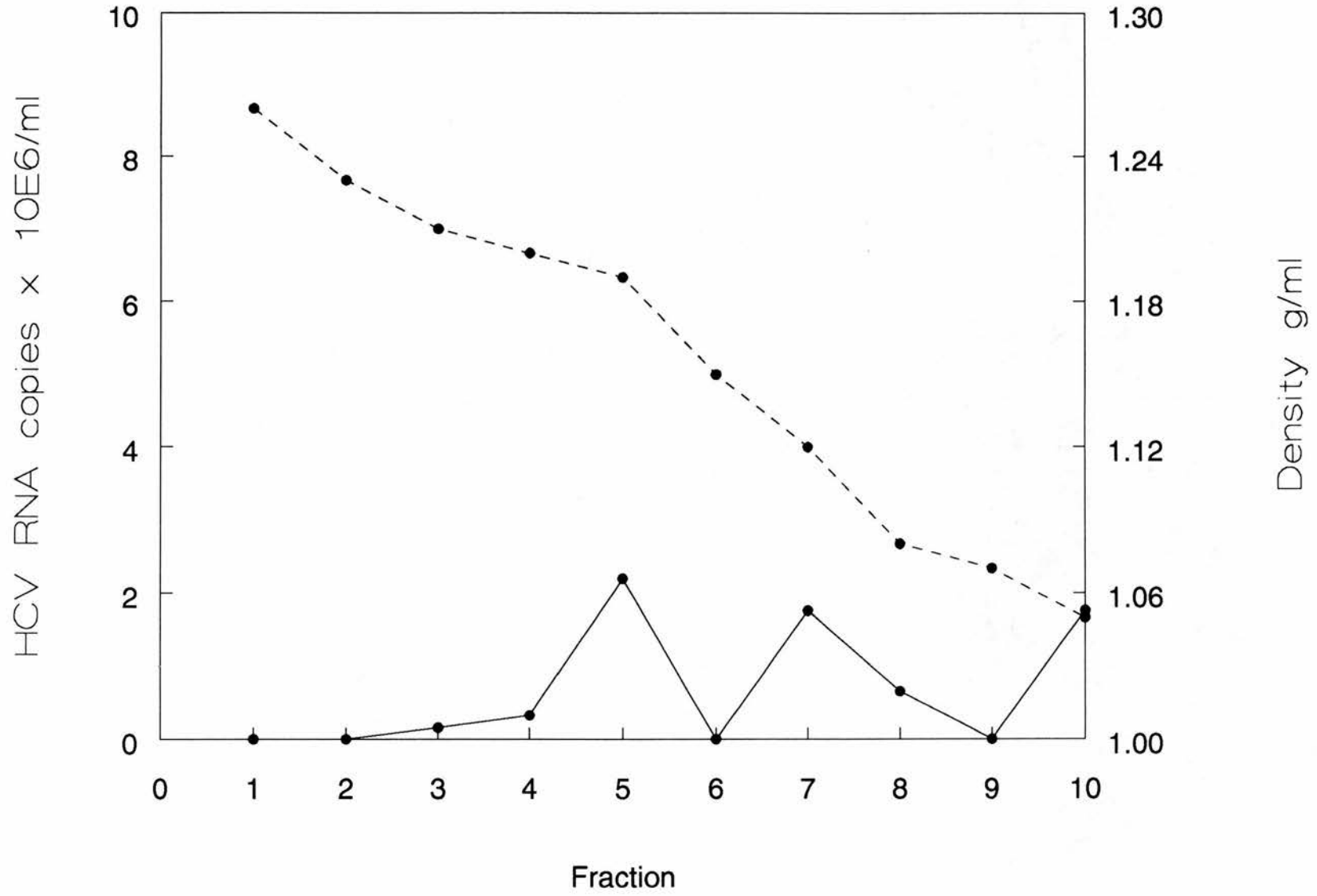


Figure 4.3 Sedimentation profile of sample C in sucrose gradient.

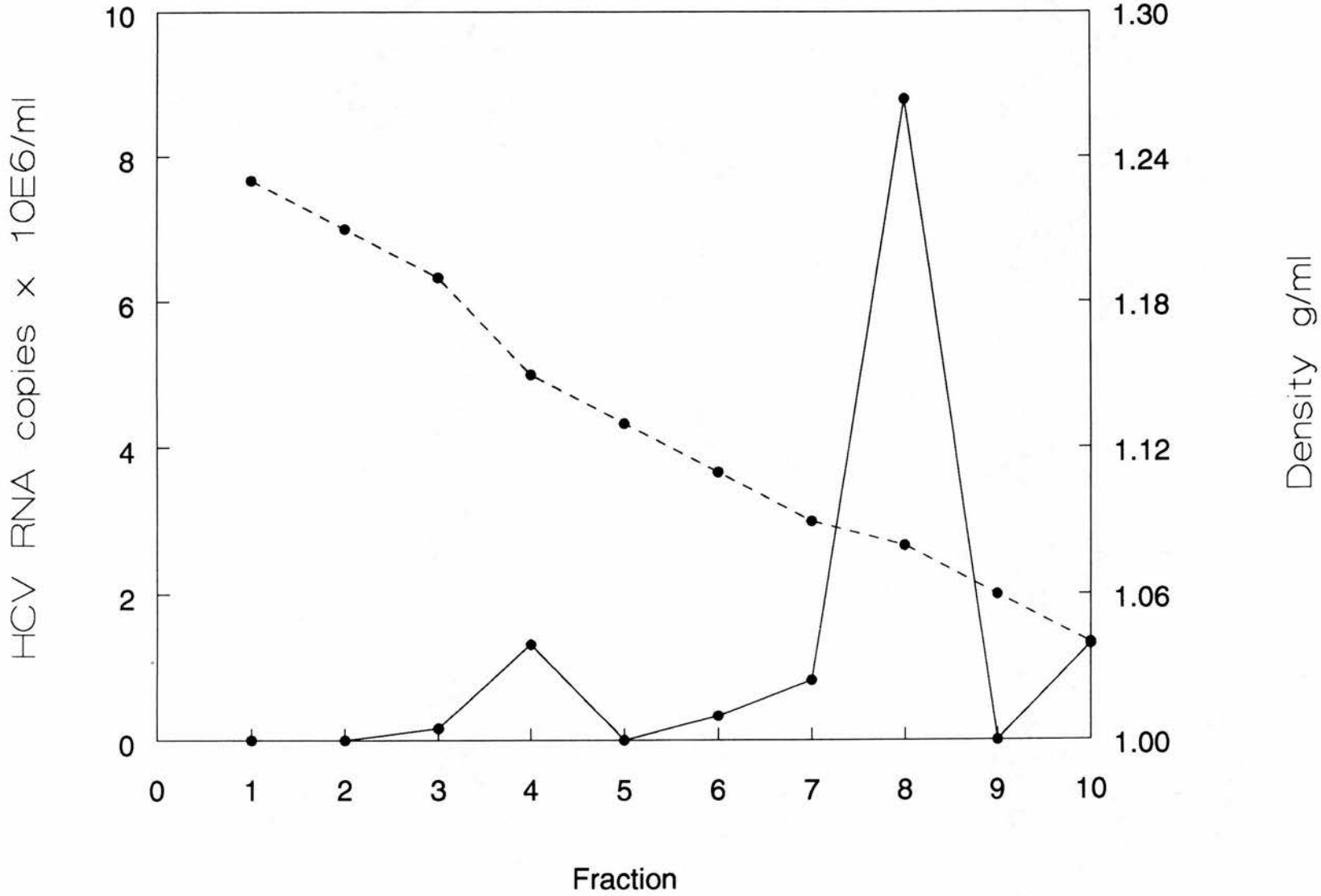


Figure 4.4 Sedimentation profile of sample F in sucrose gradient.

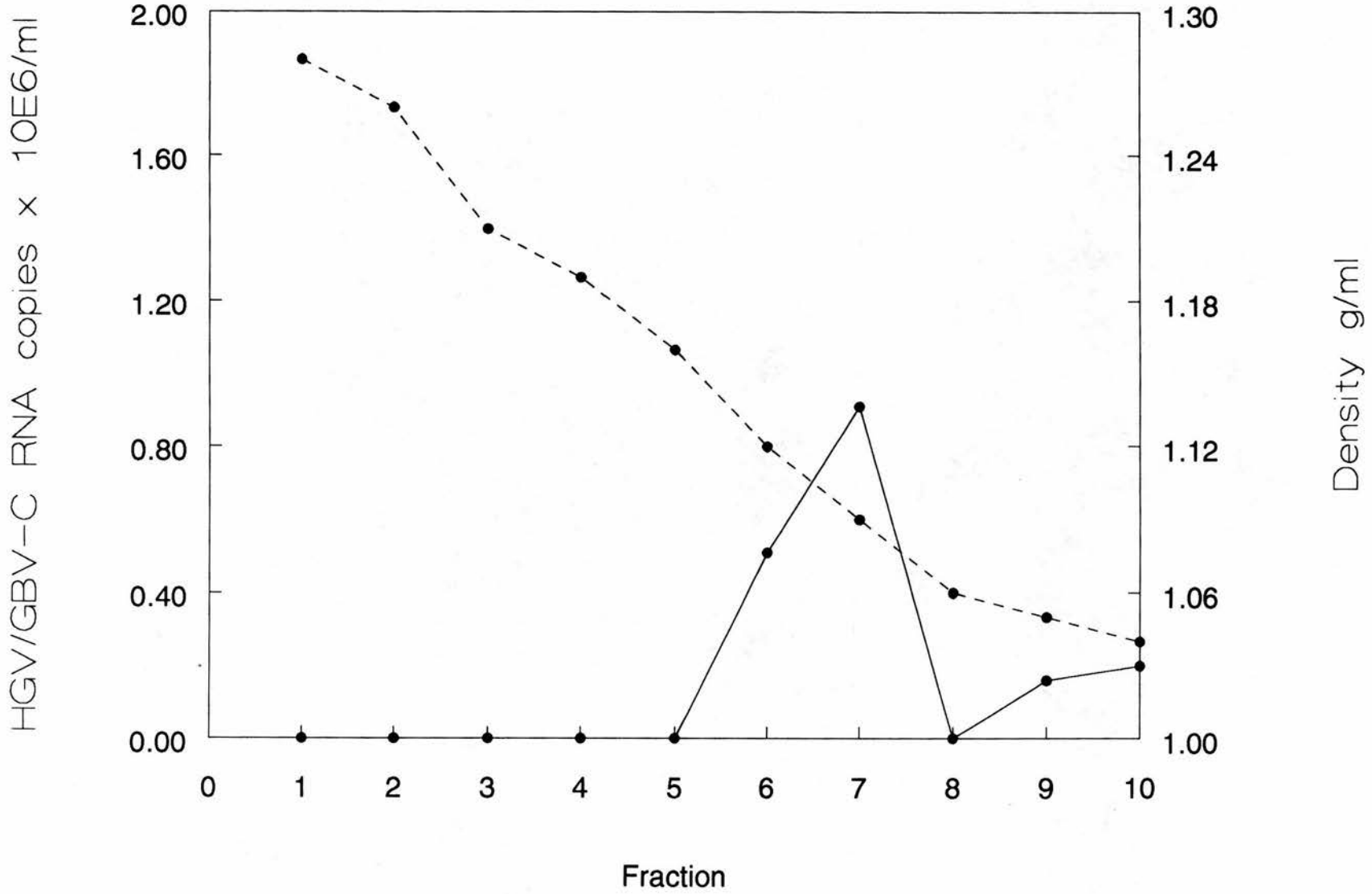


Figure 4.5 Sedimentation profile of sample G in sucrose gradient.

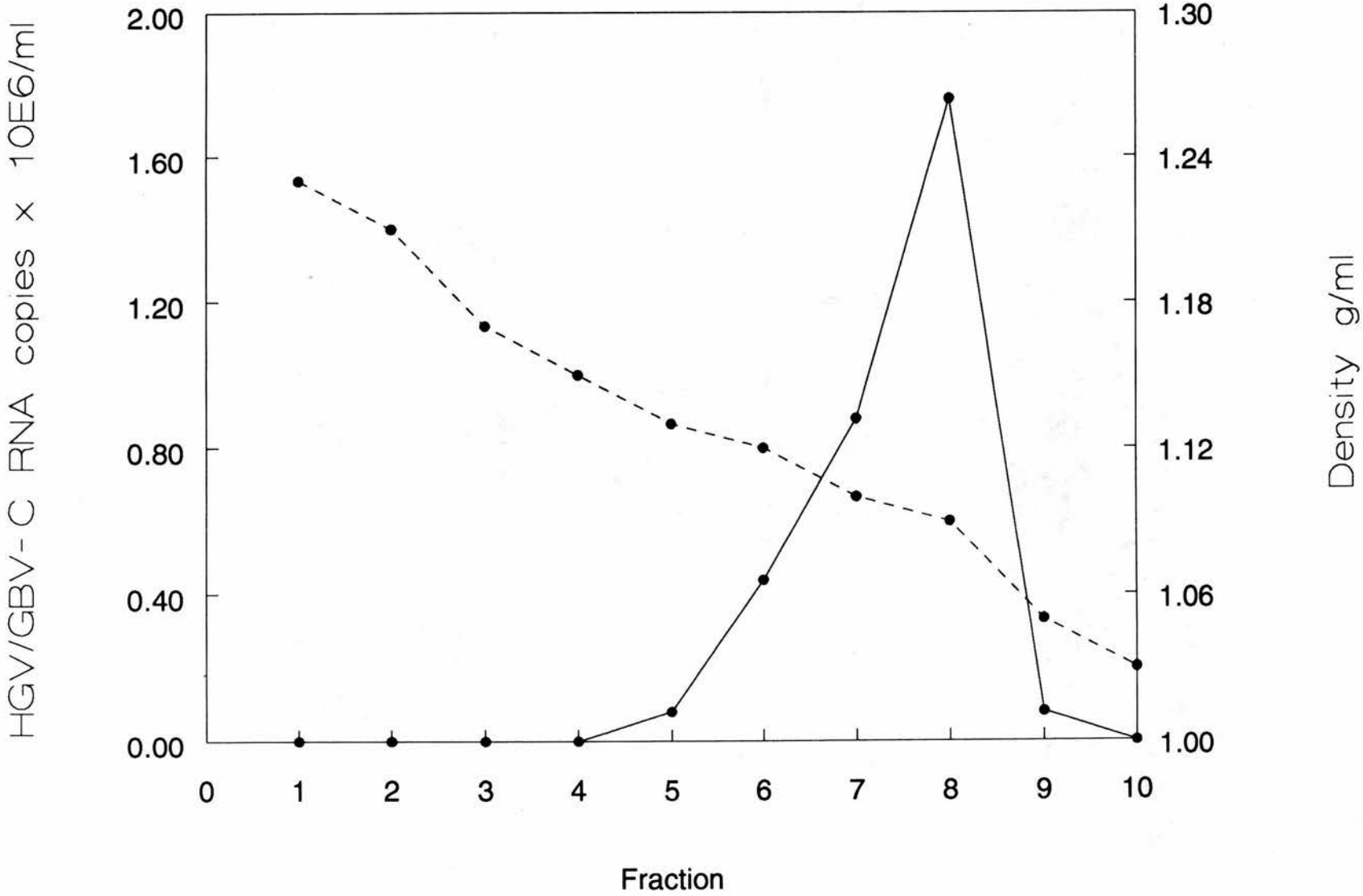
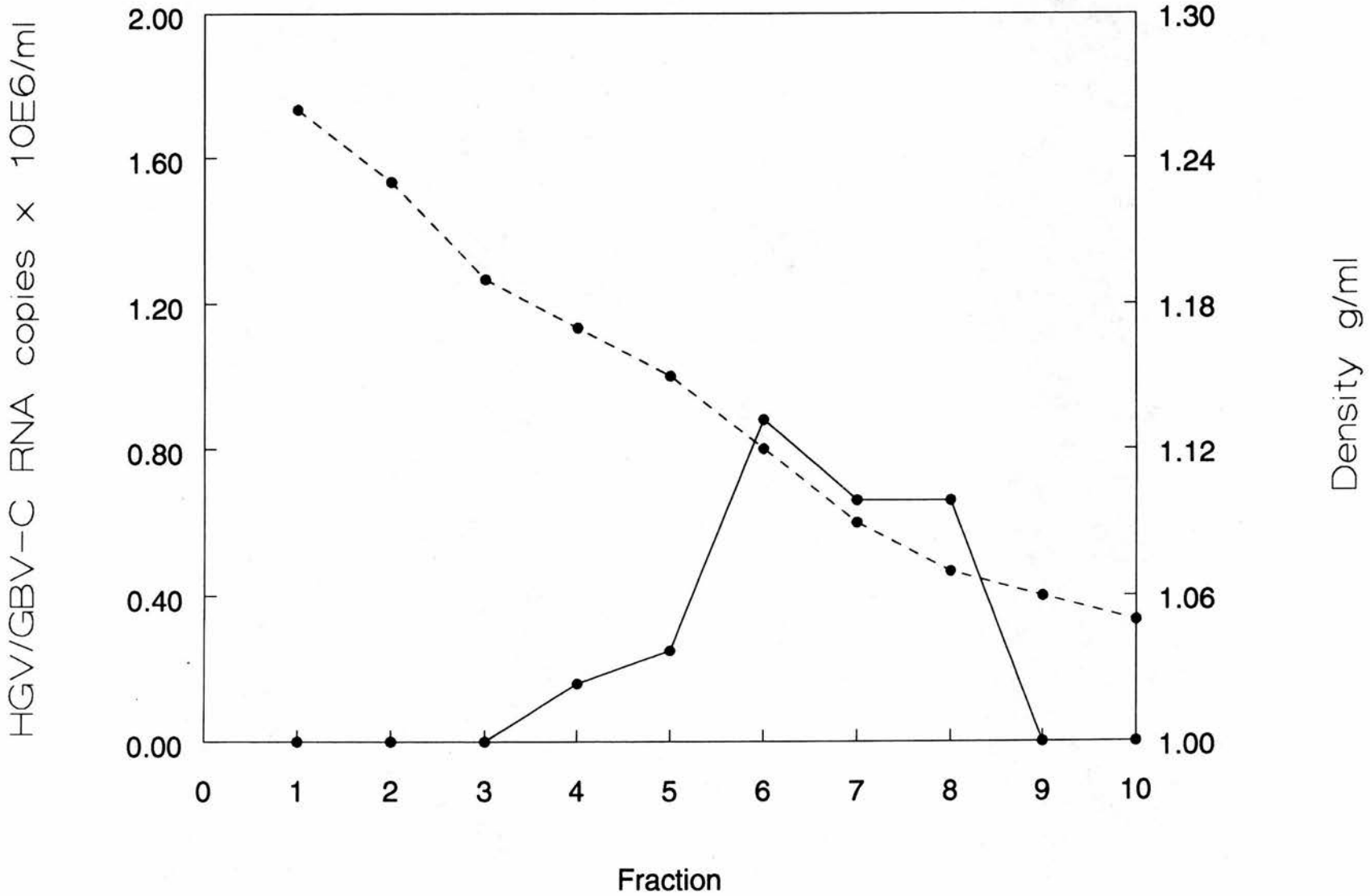


Figure 4.6 Sedimentation profile of sample H in sucrose gradient.



pellets were resuspended in 50 μl PBS and then 10 μl of each sample was used to prepare negatively stained grids for electron microscopy examination following a routine procedure: the suspension was applied to formvar-carbon grids and then negatively stained with 2% phosphotungstic acid pH 6.2 for 2 min (the grids were prepared by D.Notman, Clinical Virology Laboratory, Medical School).

Virus-like spherical particles were observed only in the 1.16g/ml fraction from sample A (HCV infected) and these closely resembled in appearance and size published photographs of HCV virions (Kaito *et al.* 1994). In the other fraction of sample A and both fractions of sample F, partially decomposed or incomplete virus-like particles were observed. The presence of lipoprotein particles (in the fractions with low density) and sucrose, the changes in the osmotic pressure and the repeated sample freezing-thawing steps during the preparation may have been damaged the integrity of the virus particles, making it difficult to visualize them.

4.3 DISCUSSION

Although HCV and HGV/GBV-C are closely related viruses, comparison of their physico-chemical properties suggest important structural differences that may be related to differences in genome organization.

HCV RNA was present in fractions with a low density in all three samples: at 1.04-1.06 g/ml for sample A, at 1.05 and 1.12 g/ml for sample B and

1.08 g/ml for sample C; these density values correspond with those previously shown to represent HCV virions associated with low-density lipoproteins (Bradley *et al.* 1991; Thomssen *et al.* 1992; Hijikata *et al.* 1993; Thomssen *et al.* 1993). Samples A and B also contained HCV RNA in fractions with higher densities (1.16 g/ml for sample A, 1.19 g/ml for sample B). This is consistent with the presence of immune complexes representing HCV particles associated with immunoglobulin (Hijikata *et al.* 1993; Thomssen *et al.* 1993; Kanto *et al.* 1995). This conclusion is supported by the observation that samples A and B were both positive by ELISA for anti-HCV antibodies and this reactivity was confirmed by the RIBA-3 assay (Table 4.1) while in sample C which was ELISA-negative, relatively little HCV RNA was detected at 1.15g/ml.

A very different sedimentation profile was observed for HGV/GBV-C in sucrose gradients. A single peak of viral RNA was present in low density fractions within a range of 1.07 to 1.12 g/ml, corresponding to the density of HCV virions associated with serum lipoproteins and/or other plasma proteins (Bradley *et al.* 1991; Thomssen *et al.* 1993). Similar results have subsequently been reported (Sato *et al.* 1996; Melvin *et al.* 1998). Direct evidence for an association between HGV/GBV-C and LDL comes from the demonstration that antibodies to human lipoproteins A-I and B could precipitate HGV/GBV-C from low density fractions (Sato *et al.* 1996). Unlike HCV, no HGV/GBV-C RNA was detected in fractions with densities higher than 1.17 g/ml expected for virus particles in immune complexes. These observations indicate that the HGV/GBV-C particles from human plasma are associated with LDL and other

plasma proteins but not bound to antibodies. This conclusion is confirmed by the observation that HGV/GBV-C could not be efficiently precipitated when plasma samples positive for both HCV and HGV/GBV-C RNA were incubated with anti-human IgG (Hijikata and Mishiro, 1996). Similarly, no HGV/GBV-C RNA could be detected in density fractions or sera incubated with goat antibodies to human IgG (kappa light chain) and precipitated with rabbit anti-goat IgG (Sato *et al.* 1996). The lack of evidence for HGV/GBV-C in immune complexes is consistent with the observation that antibodies to the E2 protein of HGV/GBV-C can only be detected in sera that are negative for HGV/GBV-C RNA, suggesting that it is the antibody response that leads to clearance of viraemia (Dille *et al.* 1997; Tacke *et al.* 1997). In contrast, in HCV infection, viraemia and antibodies specific for HCV coexist, with 97.3% of HCV-RNA positive sera shown to contain antibodies to HCV-E2 (Lesniewski *et al.* 1995). At the same time, attempts to detect other specific antibodies to HGV/GBV-C were unsuccessful. Although immunoreactive epitopes were found to be located in regions of the HGV/GBV-C putative NS3, NS4 and NS5 nonstructural proteins (PilotMatias *et al.* 1996), antibodies to prokaryotically expressed HGV/GBV-C recombinant proteins were infrequently detected amongst individuals that were HGV/GBV-C RNA positive by PCR (Dawson *et al.* 1996; PilotMatias *et al.* 1996).

No HGV/GBV-C RNA was detected in fractions with densities higher than 1.17 g/ml, whereas 1.21 g/ml fractions were PCR-positive for HCV gradients (Table 4.2), consistent with the presence of nucleocapsids (Thomssen

et al. 1992). In another published sucrose density study, treatment of HGV/GBV-C positive plasma samples with 1% chloroform before fractionation did not lead to the appearance of HGV/GBV-C RNA in higher density fractions, although the viral titre was substantially reduced following treatment with chloroform at higher concentrations (5% and 20%) (Melvin *et al.* 1998).

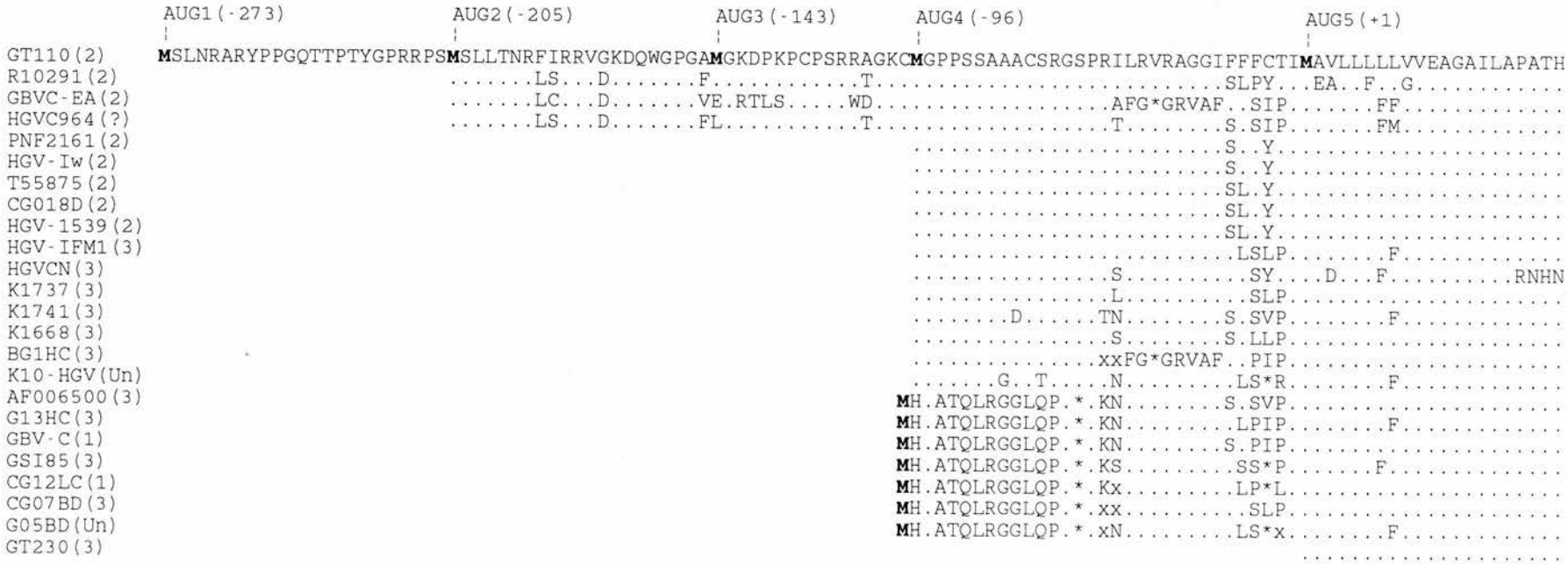
In contrast, another recent study reported that HGV/GBV-C RNA was present in sucrose fractions with densities of 1.07 and 1.17 g/ml while after a two-step extraction of the plasma samples with an equal volume of chloroform, viral RNA was only present in fractions with densities of 1.22 to 1.24 g/ml (Xiang *et al.* 1998). A similar shift in the HGV/GBV-C RNA peak to fractions with heavier density (1.23-1.26 g/ml) (Sato *et al.* 1996) or (1.25-1.29 g/ml) (Melvin *et al.* 1998) occurred following the treatment of the plasma samples with 5% Tween-80 and RNasin. However, after this treatment, the RNA could not be passed through 200 nm filter suggesting that it was converted into an extended form since particles of the expected size of nucleocapsids (diameter of 20-30 nm) should pass through filters of this pore size (Melvin *et al.* 1998). Detection of HGV/GBV-C RNA was significantly reduced or lost after treatment with Triton X-100 or NP40 in a concentration of 0.5% or higher (Sato *et al.* 1996; Melvin *et al.* 1998). On the other hand, detergent treatment at concentrations that did not reduce the viral titre produced no change in the density of HGV/GBV-C RNA containing fractions.

These observations suggest that HGV/GBV-C lacks a nucleocapsid. This conclusion is consistent with the sequence analysis of several HGV/GBV-C

complete genome sequences in which no conserved open reading frame exists upstream of the E1 protein coding region that is capable of encoding a core-like protein (Erker *et al.* 1996; Okamoto *et al.* 1997; Wang *et al.* 1997).

In order to investigate this question more completely, a detailed comparison was made of this region of the HGV/GBV-C genome in the twenty-four complete genome sequences available in GenBank database in May 1998. The 5'-terminal fragments (618 nucleotides long) were aligned and then translated in three reading frames using Simmonic 2000 sequence editor program. The nucleotide alignment showed that many nucleotide insertions, deletions and substitutions were present within this region of the genome. Although there are up to five AUG (methionine Met) codons in-frame with the HGV/GBV-C polyprotein, the only AUG codon conserved amongst all twenty-four sequences is that of the NH_2-terminus of the E1 gene (AUG-5) (Figure 4.7). This codon was identified as the site of translation initiation in a GBV-C isolate in a cell-free in vitro translation system (Simons *et al.* 1996). Comparison of amino acid sequences preceding this methionine codon showed that the open reading frames (ORFs) of the HGV/GBV-C isolates differ significantly in composition and length even within the same phylogenetic grouping (Figure 4.7). As a result, the ORF starts from AUG-1 for 1 out of 24 isolates (1/24), from AUG-2 for 3/24, from AUG-4 for 12/24, from AUG-5 for 1/24 and 7/24 have the protein encoded from an additional AUG (Figure 4.7). For example, GT110 isolate contains the longest sequence of 91 amino acids. In two isolates - R10291 and HGVC 964 - the ORF containing 68 amino acids starts at the second AUG

Figure 4.7 Amino acid sequence alignment of the putative core region of HGV/GBV-C complete genome sequences. The phylogenetic grouping of each isolate based upon analysis of complete genome sequences is indicated in brackets. Sequence identities are indicated by dots. Methionine codons (M) in-frame with that of the polyprotein are shown in bold and numbered (AUG1-4) relative to their position 5' to the M codon (AUG5) previously identified as the NH2 terminal of the virus polyprotein (Simons,J.N. *et al.*, 1996). The putative signal sequence of E1 is underlined. X=undefined codon because of ambiguities in nucleotide sequence; * = stop codon.



codon (position -205) that is in-frame with the conserved one. The starting codon for GBVC-EA amino acid sequence would also be the second AUG but the presence of an in-frame stop codon at position -37 disrupts the sequence. In some isolates (PNF2161, HGV-Iw, T55875, CG018D, HGV1539, HGV-IFM1), the fourth AUG codon at position -96 represents the start of their ORF of 31 amino acids. In BG1HC isolate, this putative amino acid sequence is disrupted due to the presence of an in-frame stop codon at position -37. The ORF of other isolates (AF006500, G13HC, GBV-C, GSI85, CG12LC, CG07BD, G05BD) starts at an additional AUG codon just upstream of the fourth one but is disrupted by a stop codon at position -55. In GT230 isolate, the ORF starts only at the fifth AUG codon. Similar conclusions were revealed from the analysis of 5'-terminal sequences amplified from the sera of HGV/GBV-C infected individuals (Muerhoff *et al.* 1996b; Pickering *et al.* 1997b).

Despite this evidence, some authors have suggested that HGV/GBV-C does encode a core protein of 91 amino acids long (Xiang *et al.* 1998). A single nucleotide insertion 43 amino acids into this putative core protein would be responsible for its lack of conservation among different HGV/GBV-C isolates. In fact, sequence analysis reveals that heterogeneity of this part of the HGV/GBV-C genome is due to the presence of nucleotide insertion or deletion (Muerhoff *et al.* 1996; Okamoto *et al.* 1997; Wang *et al.* 1997) so the possibility of consistent sequencing errors is excluded. The amino acid sequence starting at the fourth AUG codon is relatively well conserved amongst a number of isolates (Figure 4.7), (Pickering *et al.* 1997b) and a fourteen amino acid-

fragment within this ORF has been used as a synthetic peptide in an ELISA to test the presence of antibody to the HGV/GBV-C core protein in HGV/GBV-C infected individuals (Xiang *et al.* 1998). Antibody to this peptide was detected in plasma samples from HGV/GBV-C RNA positive patients but not in negative controls, a result that could indicate that this peptide sequence is expressed *in vivo*. However, since only four samples were tested in ELISA and the absorbance values were relatively low, the generality of these findings is unknown.

Recently, the ability of the HGV/GBV-C anti-genome sequence to encode a core-like protein was investigated (Kondo *et al.* 1998). The analysis of the anti-genome of thirty-eight HGV/GBV-C isolates led to the identification of a consensus sequence of 118 amino acids, in which the methionine and the stop codons were conserved amongst all isolates. Overlapping synthetic peptides based on this fragment located within the E2 region of the HGV/GBV-C anti-genome were prepared and then used for the detection of specific antibodies by ELISA. Although 66.7% of the HGV/GBV-C RNA positive sera tested were positive for some peptides, false-positive reactions occurred and the sensitivity of the assay was relatively low. Further studies are needed in order to assess the existence of a new ORF within the HGV/GBV-C anti-genome and the possibility of the ambisense protein to function as core protein.

Viral core proteins protect the viral genome from the RNases and enable the encapsidation, and so the formation of a nucleocapsid is considered essential for many viruses. For example, the HCV core protein is capable of

multimerization *in vitro* and *in vivo*, in a process that may represent the first step of virus assembly (Matsumoto *et al.* 1996). Kinetic studies of glutaraldehyde cross-linking of HCV core protein expressed in mammalian cells revealed that the region responsible for multimerization lies within the amino-terminal hydrophilic domain (amino acids 1-115). Whether the mechanism of HGV/GBV-C virion assembly takes place in the absence of a nucleocapsid or by using other virus or host specific proteins with biological properties similar to those of a virus core protein has yet to be discovered.

In conclusion, the buoyant density of HGV/GBV-C virus particles in human sera is estimated to be between 1.07 - 1.12 g/ml, much lower than that of the other members of *Flaviviridae* family, except HCV. This low density seems to be related to the binding of viral particles to human plasma components, such as LDL. The distinct sedimentation behaviour of HGV/GBV-C in sucrose gradient from that of HCV may reflect structural differences between the HCV and HGV/GBV-C genomes, consistent with HGV/GBV-C lacking a core gene.

CHAPTER 5

5. ANALYSIS OF THE PRIMARY AND SECONDARY STRUCTURE OF THE HGV/GBV-C 3' UNTRANSLATED REGION

5.1 INTRODUCTION

It has been proposed that the termini of an RNA virus genome present certain nucleotide sequences functionally important for viral RNA replication, transcription, translation and packaging (Strauss and Strauss, 1983). These elements could be represented by conserved nucleotide sequences that are recognized by viral enzymes or these sequences could be part of secondary structures stabilised by hydrogen bonds in which the structure, rather than the sequence, is recognized. The genomic organization of HGV/GBV-C is similar to that of other members of *Flaviviridae* family, with a long ORF flanked by a 5'-UTR and a 3'-UTR but at present there is very little information about the secondary structure of these regions or their role in virus replication.

Amongst the *Flaviviridae*, the primary and secondary structures of the 5'- and 3'-UTR have been described for flaviviruses (Brinton *et al.* 1986; Hahn *et al.* 1987; Proutski *et al.* 1997), pestiviruses (Deng and Brock, 1993) hepatitis C viruses (Brown *et al.* 1992; Smith *et al.* 1995; Kolykhalov *et al.* 1996; Blight and Rice, 1997) and GB viruses (Simons *et al.* 1996). Certain structural elements in these regions are functionally important for viral translation, replication, transcription and assembly, although the precise mechanisms of these events are not well understood. For example, the 3'-UTR of the flavivirus Yellow Fever

virus (YFV) contains a set of three closely spaced repeated sequences, each approximately 40 nucleotides long, while a stable secondary structure is formed by the 3'-terminal 87 nucleotides (Rice *et al.* 1985). Comparison of the 3'-terminal sequences of West Nile virus (WNV), Saint Louis Encephalitis (SLE) and YFV has identified features common amongst different flaviviruses (Brinton *et al.* 1986). The existence of a stable 3'-terminal secondary structure was confirmed by direct RNA sequencing and partial enzymatic cleavage methods (3'-end radiolabeling of genomic RNA followed by partial digestion with specific RNases and electrophoresis of the products on sequencing gels). Seven regions of conserved sequence of variable lengths were identified among the 3'-UTRs of WNV, SLE and YFV and these may represent signals for viral polymerase recognition and binding. The conservation in size and structure of this region supports the hypothesis that it contains important elements for viral RNA replication.

Besides these subgroup-specific features, two short conserved RNA sequences (CS1 and CS2) , located 5' to the 3'-terminal structure, are shared by all mosquito-borne flaviviruses and are considered to play important cis-acting roles in RNA replication or packaging (Hahn *et al.* 1987). CS1 is about 26 nucleotides in length and located immediately adjacent to the terminal secondary structure. Eight contiguous nucleotides of CS1 were found to be complementary to an eight-nucleotide conserved domain in the 5'-region of the flavivirus RNA. Their base-pairing could lead to cyclization of the viral genome that may be important for regulating genome translation, replication or packaging. CS2 is

approximately 24 nucleotides in length, highly conserved, and located 12 to 22 nucleotides upstream of CS1. CS2 is present in two copies in members of the Dengue (DEN) and Japanese Encephalitis (JE) subgroups.

Imperfectly repeated sequences have also been identified in the 3'-UTR of pestiviruses, although the nucleotide sequence and location of the repeats is not conserved between different viruses (Deng and Brock, 1993). A secondary structure model for the 3'-UTR of pestiviruses contains a long conserved 3'-terminal stem-loop (Deng and Brock, 1993).

Similar features were reported for HCV, which is classified with flaviviruses and pestiviruses in the *Flaviviridae* family. The viral genome has a 3'-UTR that consists of three domains: a highly variable sequence (21 to 40 nucleotides) followed by a poly U/polypyrimidine tract of variable length and composition and a 98 nucleotide-highly conserved sequence (Kolykhalov *et al.* 1996). Computer prediction of the secondary structure and enzymatic analysis are consistent with this region having a stable and conserved stem-loop (SL) within the 3'-terminal 46 bases suggesting that this element may represent a recognition site for viral and/or cellular proteins (Kolykhalov *et al.* 1996; Blight and Rice, 1997). A 58-kDa cellular protein identified as the polypyrimidine tract-binding protein (PTB) binds to the highly conserved 3'-terminal 98 nucleotides of the HCV RNA (Ito and Lai, 1997); protein binding seems to require secondary structure elements and specific sequences within the 98 nucleotide-region.

In contrast, the 52 nucleotides at the 5'-end of the conserved 98 nucleotide 3'-fragment do not seem to form a stable stem-loop structure. However, this conserved sequence may play a role in the recognition of viral RNA by proteins in a sequence-specific manner. A significant fraction of HCV genome RNA from a highly infectious inoculum contains the 3'-terminal 46 base-element which is likely to be essential for authentic HCV replication and recovery of infectious RNA from cDNA. Indeed, a recent study reported that this region of the 3'-UTR of the HCV is essential for infectivity *in vivo* (Yanagi *et al.* 1998).

For HGV/GBV-C, there have been very few reports of the sequence or secondary structure of the 3'-UTR of different isolates (Erker *et al.* 1996; Nakao *et al.* 1997; Okamoto *et al.* 1997; Katayama *et al.* 1998). The secondary structure model of Okamoto *et al.* was constructed only for the 3'-terminal 140 nucleotides of the 3'-UTR and based on only three sequences (Okamoto *et al.* 1997) while Katayama *et al.* only predicted the structure of the last 30 nucleotides of 3'-UTR (Katayama *et al.* 1998).

This chapter describes a more comprehensive comparison and analysis of the primary sequence and secondary structure of the 3'-UTR. Nucleotide sequences were obtained for different HGV/GBV-C isolates from Edinburgh, Pakistan, Zaïre, Brazil, Gambia and Papua-New Guinea. These sequences were compared with published sequences and used to construct a consensus secondary structure model for the whole 3'-UTR of the HGV/GBV-C genome.

5.2 RESULTS

PRIMARY STRUCTURE

Twenty-two partial or complete sequences of the 3'-UTR of HGV/GBV-C derived from complete genome isolates were obtained from GenBank in May 1998 (Table 5.1). Seventeen additional nucleotide sequences were obtained by sequence analysis of PCR products amplified with primers directed against conserved regions of HGV/GBV-C at the carboxyl-end of the NS5B gene and the extreme 3'-end, respectively (Table 2.1). The sera analysed were those used for sequence analysis of the 5'-UTR (Edinburgh, Pakistan and Zaïre samples) (Chapter 3) as well as others from Brazil, Papua-New Guinea and Gambia samples identified as being PCR positive for the 5'-UTR of HGV/GBV-C (Prescott,L.-unpublished). PCR products were cloned into pGemT vector (section 2.4.2) and sequenced (section 2.3.3).

All of the sequences were collinear with only two sites of deletion/insertion (positions 8705 and 8709). The alignment revealed highly conserved regions between positions 8530 (UGA stop codon) to 8558, 8580 to 8675, 8680 to 8703, and 8715 to 8844, while the intervening regions showed limited variability (Figure 5.1). The high degree of sequence conservation in this region is consistent with the existence of important recognition signals for the virus-specific replicase (Strauss and Strauss, 1983).

Comparison of the thirty-nine HGV/GBV-C 3'-UTR sequences offered the possibility of searching for putative repeat sequences similar to those

Table 5.1 HGV/GBV-C isolates used in 3'UTR analysis. The sequences obtained in this study are identified by prefixes Pak for Pakistan, Ed for Edinburgh, Zai for Zaire, Brz for Brazil, Png for Papua-New Guinea and Gam for Gambia.

Sequence name	Source	Accession number	Group
GT110	Genbank	D90600	2
HGV-1539	Genbank	AF031829	2
T55875	Genbank	AF031827	2
HGV-1517	Genbank	AF031828	2
PNF2161	Genbank	U44402	2
HGV-Iw	Genbank	D87255	2
GBVC-EA	Genbank	U63715	2
Ed_34	this study	-	2
Ed_62	this study	-	2
Ed_45	this study	-	2
Ed_69	this study	-	2
Brz_I	this study	-	2
Pak_502	this study	-	2
Pak_540	this study	-	2
Pak_566	this study	-	2
HGVCN	Genbank	U94695	3
GSI85	Genbank	D87262	3
GSI93	Genbank	D87263	3
Ed_81	this study	-	3
K1737	Genbank	D87709	3
K2141	Genbank	D87713	3

Table 5.1 continued.

K1741	Genbank	D87710	3
K1789	Genbank	D87711	3
GT230	Genbank	D90601	3
HGV-IFM1	Genbank	AB008342	3
K606	Genbank	D87708	3
K1668	Genbank	D87714	3
K3732	Genbank	AB008335	3
K1775	Genbank	D87715	3
K1916	Genbank	D87712	3
GBV-C	Genbank	U36380	1
Zai_8	this study	-	1
Zai_9	this study	-	1
Zai_11	this study	-	1
Brz_25	this study	-	1
Png_29	this study	-	not known
Png_48	this study	-	not known
Gam_6	this study	-	not known
Gam_2	this study	-	not known

Figure 5.1 Alignment of HGV/GBV-C 3'-UTR sequences. The entire 3'-UTR (315 nucleotides) as well as an extra 20 nucleotide-fragment upstream of the polyprotein termination codon UGA is shown. Sequence identity with GT110 is indicated by dots and the missing information by spaces. The repeated and conserved sequences are underlined; - = gap introduced to preserve alignment.

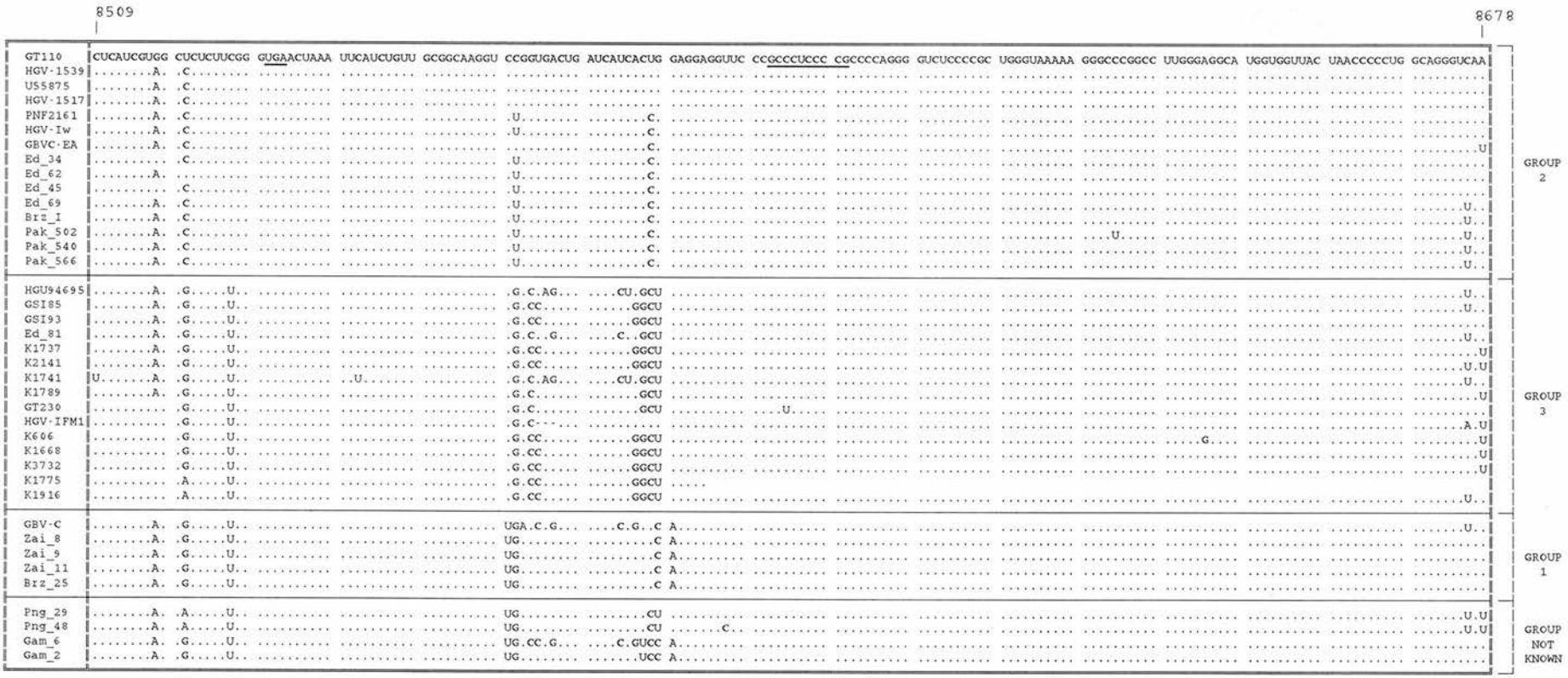
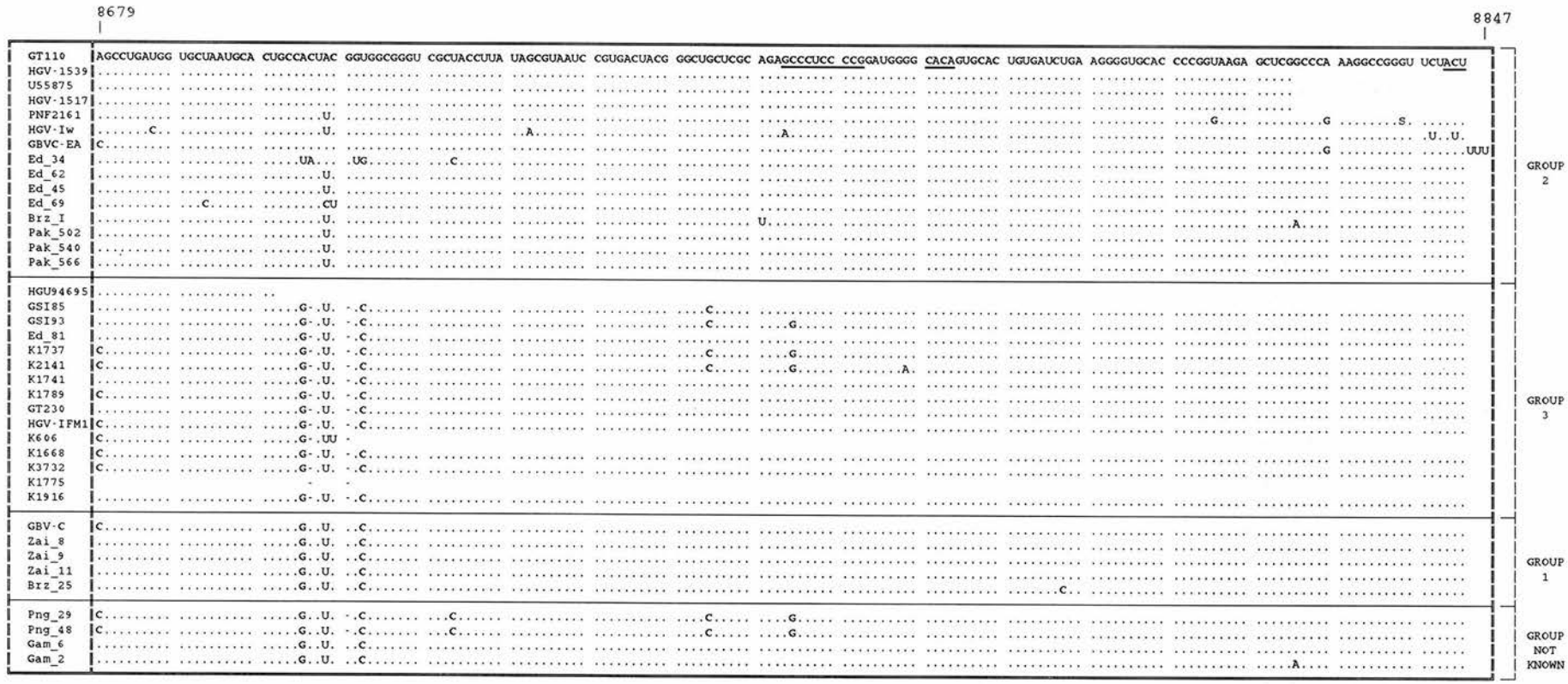


Figure 5.1 continued

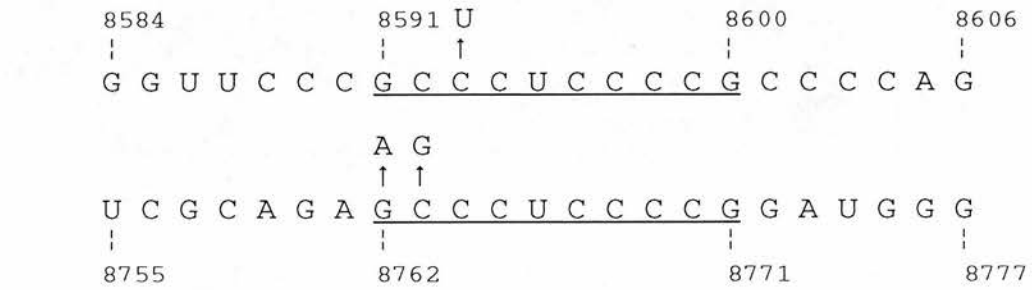


identified in flaviviruses and pestiviruses that may represent viral replication signals (Rice *et al.* 1985; Deng and Brock, 1993). In this study, the search for the nucleotide sequence repetition was done using fragments of ten nucleotides each that were aligned by Simmonic 2000 sequence editor program. It revealed only one repeat sequence of 10 nucleotides located at positions 8591 and 8762, respectively, that was conserved among all the viral isolates with few exceptions. Substitutions occurred at position 8593 (C→U) for GT230, at position 8763 (C→G) for five isolates (GSI93, K1737, K2141, Png_29, Png_48) and at position 8762 (G→A) for HGV-Iw (Figure 5.2 A). In addition, two short, highly conserved sequences were identified: CACA located at position 8779 and ACU at position 8842 (the last three nucleotides of the 3'-UTR fragment) (Figure 5.1). The first sequence is also present in the 3'-terminal region of flavivirus (Brinton *et al.* 1986) and pestivirus genomes (Deng and Brock, 1993); the second sequence occurs at the extreme 3'-end of flavivirus genome RNAs (Brinton *et al.* 1986). The conservation of these elements suggests that they are important in virus replication.

Complementarity between the 5'-UTR and 3'-UTR of the viral genome of positive strand RNA viruses has been described previously, consistent with the viral replicase having similar recognition sites for (+) and (-) strand synthesis (Strauss and Strauss, 1983; Rice *et al.* 1985). However, the 5'- and 3'-terminal sequences of HGV/GBV-C genome are not complementary (Figure 5.2 B). A six nucleotide-sequence in the 3'-UTR at positions 8724-8729 is complementary to the positions -293 - -288 of the 5'-UTR (Figure 5.2 C). This was the longest

Figure 5.2 Primary structure of HGV/GBV-C 3'-UTR

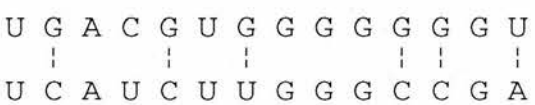
A. Repeated sequence in the 3'-UTR of HGV/GBV-C RNA.The identical nucleotides are underlined. Substitutions are indicated by arrows.



B. 5'- and 3'-terminal sequences of HGV/GBV-C RNA

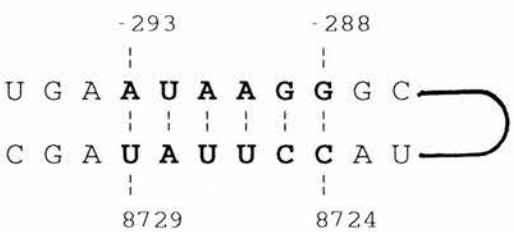
(+) strand

5'-end



3'-end

C. Nucleotide complementarity between 5'-UTR and 3'-UTR of HGV/GBV-C RNA. The complementary sequences are shown in bold.



complementary sequence found by manual search.

The nucleotide composition of the 3'-UTR of HGV/GBV-C (calculated using the MEGA package (Kumar *et al.* 1993)) is 17.5% A, 30.1% C, 31.3% G, 21.1% U with a GC content of 61.4%, similar to that of the whole virus genome (17.88% A, 27.36% C, 31.97% G, 22.74% U). In contrast, the flavivirus genome is characterized by a lower GC content and a relatively high purine content (27.3% A, 21.3% C, 28.4% G, 23% U) (Rice *et al.* 1985).

SECONDARY STRUCTURE

Secondary structure prediction of the 3'-UTR of HGV/GBV-C genome was performed using the RNAdraw program (Matzura and Wennborg, 1996) for 2 or 3 different sequences from each of the four phylogenetic groups defined by the analysis of complete genome sequences (Chapter 3). G-U base pairs were accepted for this analysis, in addition to Watson-Crick base pairs and the optimal structure/basepair-probability matrix/heat curve calculation algorithms in RNAdraw were imported from the Vienna RNA package (Hofacker *et al.* 1994).

Comparison of secondary structures predicted for examples of the different phylogenetic groups revealed several hairpins that were invariant in size and position. Free energy values for the 3'-UTR structures (positions 8509 to 8847) ranged from -103.19 kcal (for the PNF2161 sequence) to -111.98 kcal (for the GBV-C sequence) (Table 5.2). Based on these similarities, a consensus

Table 5.2 Free energy values calculated at 37^0C for the predicted secondary structures of the sense strand RNA (ΔG^a) and reverse complement strand RNA (ΔG^b) for the 3'-UTR of different HGV/GBV-C isolates. The percentage reduction in free energy was calculated as mean value based on nucleotide order randomization using RNAdraw program.

Isolate	Group	ΔG^a (kcal)	% reduction	ΔG^b (kcal)	% reduction
GBV-C	1	-111.98	22.50	-105.20	21.70
Zai_8	1	-107.93	17.25	-91.91	9.06
GT110	2	-110.29	21.01	-93.22	11.0
Ed_69	2	-105.94	15.73	-96.78	12.97
PNF2161	2	-103.19	16.58	-102.58	16.11
Ed_81	3	-108.81	24.65	-94.97	9.12
K1741	3	-104.45	15.87	-90.94	14.88
Png_29	not known	-108.99	21.39	-95.56	15.09
Mean value		-107.703	19.37	-96.401	13.74

model of the secondary structure of the 3'-UTR of the HGV/GBV-C RNA was constructed containing a series of stem-loop structures numbered **I** to **VII** (Figure 5.3). Covariant nucleotide substitutions revealed by sequence alignment were then identified by manual search through the potential hairpin structures.

The extended stem-loop structure **I**, which contains 20 nucleotides of the coding sequence upstream from the stop codon UGA, was present in all phylogenetic groups. While the **Ia** region showed some variation in length (the stem is one pair-shorter in group 2 sequences), stem-loop **Ib**, located between positions 8529 and 8552, was invariant in shape and size.

Hairpin **IIa** located between positions 8558 to 8580 was highly conserved and supported by the presence of numerous covariant substitutions (Figure 5.4). Although the primary nucleotide sequence within this region varied amongst different isolates (Figure 5.1), the size and position of the hairpin remained invariant. These substitutions not only maintained the proposed base pairings but were associated with different HGV/GBV-C phylogenetic groups. Stem-loops **IIb**, **IIIa** and **IIIb** were also stable structures, formed by regions with conserved nucleotide sequence.

The nucleotide fragment between positions 8638 and 8687 was conserved, except for substitutions at positions 8676 (C→U), 8678 (A→U) and 8679 (A→C) (Figure 5.1); however, no consistent secondary structure was observed for different HGV/GBV-C isolates.

The region between positions 8688 and 8745 was predicted to form an extended stable stem-loop structure of three hairpins **IVa**, **IVb**, **IVc,** all of which

Figure 5.3 Consensus model of the secondary structure of HGV/GBV-C 3'-UTR. Positions at which covariant nucleotides were observed are shown in bold. Stop codon is underlined.

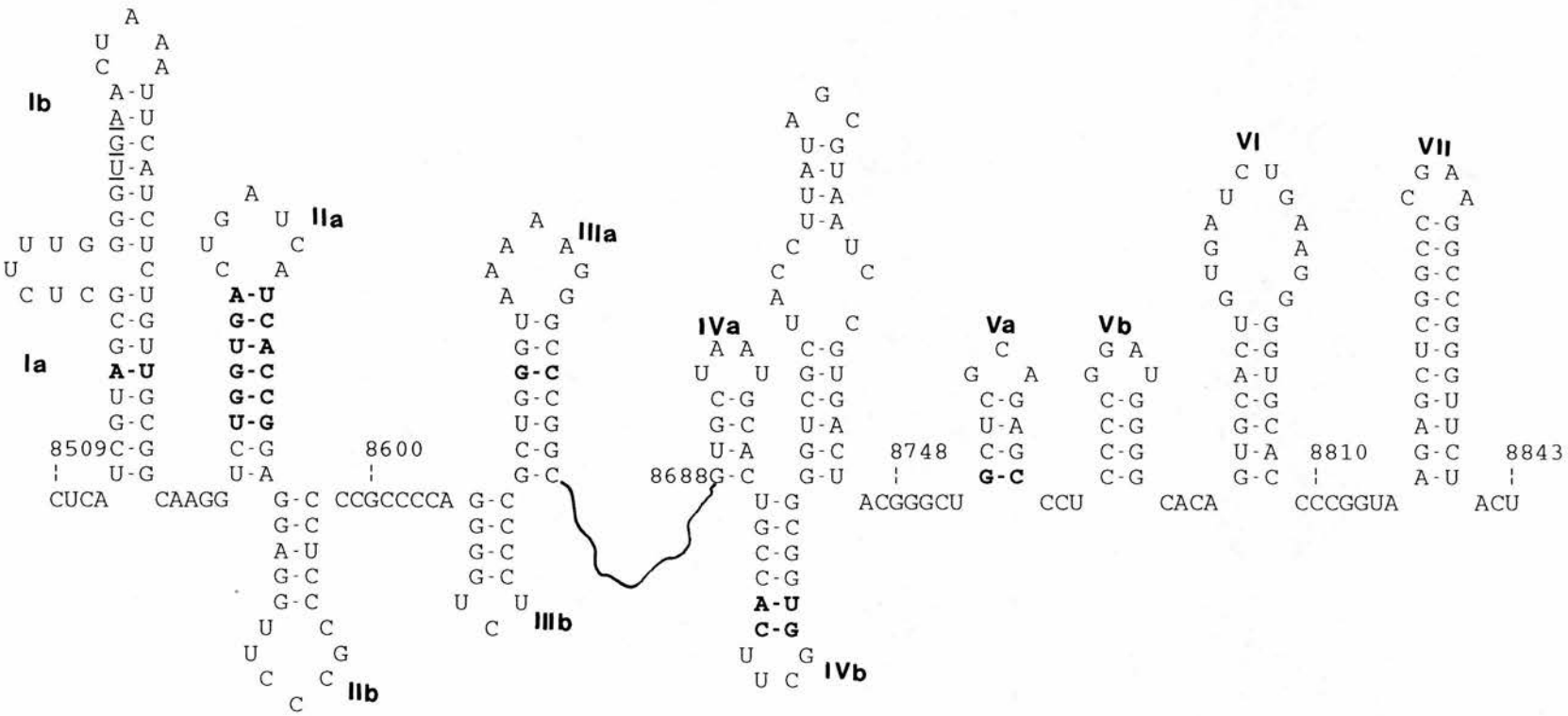
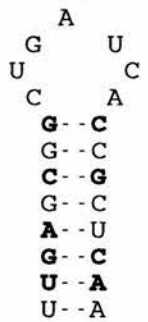
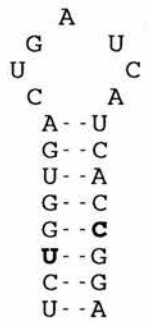


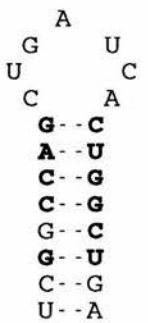
Figure 5.4 Group-specific nucleotide variation within hairpin **IIa**. Positions at which substitutions are covariant such that base pairing is maintained are shown in bold.



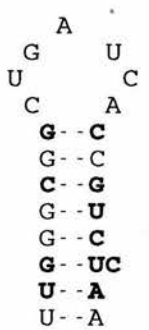
GROUP 1



GROUP 2



GROUP 3



GROUP NOT KNOWN

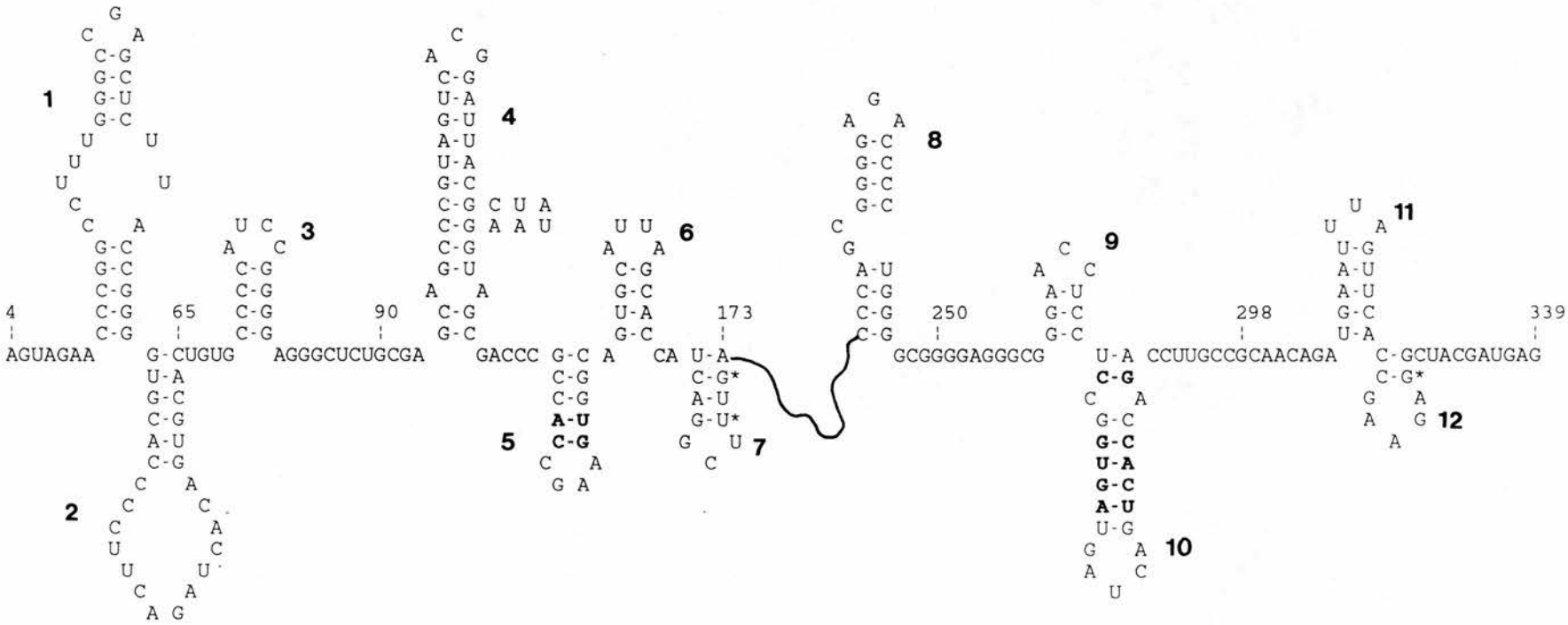
were present amongst all the sequences examined. Hairpin **IVb** contains covariant substitutions at positions 8705/8710 and 8704/87011 although the stem is one pair shorter in group 3 sequences because of deletion of positions 8705 and 8709.

The terminal part of the 3'-UTR is also conserved with stable, stem-loops - **Va** (positions 8753 and 8763), **Vb** (positions 8767 and 8778) and **VI** (positions 8783 and 8808). The few substitutions occurring within hairpin **Va** either maintained the base pairing (positions 8753/8763) or occurred in an unpaired region (position 8759 in the loop). The base-pairing of the stem is only disrupted at positions 8754/8762 (hairpin **Va**) in HGV-Iw isolate and at positions 8768/8777 (hairpin **Vb**) in K2141 isolate.

The last and longest predicted stem-loop (**VII**) is located between positions 8816 and 8841 and is well conserved and stable among all HGV/GBV-C isolates. In two cases, (Brz_I and Gam_2 sequences), the stem shape is slightly altered with its upper part disrupted by the formation of a small bulge due to a substitution at position 8824 (G→A).

A second sequence alignment representing the reverse complement (5'-terminal sequence of the minus-strand RNA) of the same HGV/GBV-C isolates was generated using Simmonic 2000 sequence editor program and a consensus model of the secondary structure was predicted using the RNAdraw program as described above. This model contains a series of stem-loop structures numbered **1** to **12** (Figure 5.5) with the free energy values for different sequences varying between -90.94 kcal (K1741 sequence) and -105.20 kcal (GBV-C sequence)

Figure 5.5 Consensus model of the secondary structure of the reverse complement strand for the HGV/GBV-C 3'-UTR. Positions at which covariant nucleotides were observed are shown in bold. Positions at which substitutions affect base-pairing are indicated by *.



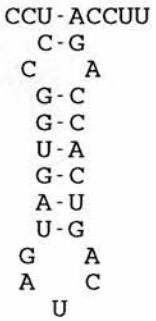
(Table 5.2).

While most of the hairpins in the minus strand were invariant in size and position for all HGV/GBV-C sequences studied, hairpins **7** (positions 163 to 173) and **12** (positions 319 to 329) were less stable. For example, the substitutions occurring within hairpin **7** at positions 170 (U→A) and 172 (G→A) would disrupt the base-pairing in all group 3 sequences apart from GSI93 and GT230 isolates and also in GBVC-EA, Ed_69, Brz_I, Pak_502, Pak_540, Pak_566 (group 2 sequences), GBV-C (group 1 sequence), Png_29 and Png_48 (unassigned group sequences).

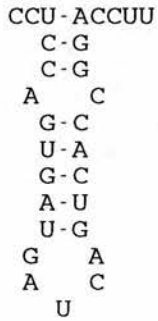
Similarly, for hairpin **12**, the base-pairing is disrupted at positions 320/328 because of G→A substitution in GT110 and Ed_62 (group 2 sequences) and G→C or G→U substitutions in group 3, group 1 and unassigned group sequences.

Hairpin **5** located between positions 134 and 148 is well conserved with covariant substitutions at positions 137/144 and 138/143, respectively. In group 3 sequences, the stem is one pair shorter because of deletions occurring at positions 139 and 142. Covariant substitutions are also present within hairpin **10** located between positions 268 and 290 although its shape is slightly different amongst the HGV/GBV-C isolates (Figure 5.6).

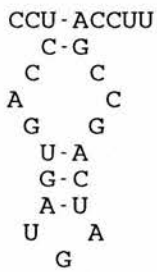
Figure 5.6 Nucleotide variation within hairpin **10**.



PNF2161, HGV-Iw, GBVC-EA, Brz_I, Pak and Ed sequences (group 2) Brz_25, Zai sequences (group 1) group 3 sequences

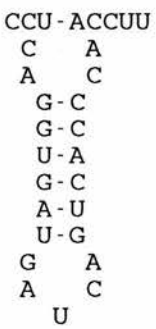


GT110, HGV-1539, T55875, HGV-1517 (group 2)



HGV-IFM1 (group 3)

GBV-C (group 1)



Png sequences

Comparative sequence analysis of the 336 3'-terminal nucleotides of the HGV/GBV-C genome from 39 isolates reveals a high degree of nucleotide sequence conservation. The existence of conserved regions could be due to the presence of structures or sequences required for viral RNA replication, translation or packaging (Strauss and Strauss, 1983). Two short sequences identified in the 3'-UTR of HGV/GBV-C genome at positions 8779 (CACA) and 8842 (ACU) respectively, are homologous to those found in 3'-UTR of flaviviruses and pestiviruses and may represent signal sequences involved in the recognition and binding of viral polymerases. A repeated sequence of 10 nucleotides found within 3'-UTR of HGV/GBV-C RNA at positions 8591 and 8762 is conserved among all isolates and may also play a role as recognition signal for viral replicase (Figure 5.2 A). At the same time, since the search for nucleotide repetition was limited to identify only perfectly repeated fragments of 10 nucleotides long, the existence of other fragments of a larger size and imperfect repeat sequence is not excluded.

The 5'- and 3'-terminal sequences of HGV/GBV-C genomic RNA are not complementary (Figure 5.2 B), implying that the recognition sites for the (+) strand and (-) strand replicase for HGV/GBV-C are different. Similarly, for pestiviruses, no complementary sequences have been found (Deng and Brock, 1993). However, the presence of a conserved six nucleotide-sequence element at position 8724 in the 3'-UTR which is complementary to a conserved sequence

at position -303 in 5'-UTR is intriguing since similar types of conserved sequences have been identified in 3'-UTR of flavivirus RNAs and are believed to represent potential cyclization sequences (Hahn *et al.* 1987; Proutski *et al.* 1997). The role of RNA cyclization in the replication process is not known; cyclization could promote the replication of full - length molecules of viral RNA if binding of the viral replicase to both 5'- and 3'-regions simultaneously was required to initiate RNA replication (Hahn *et al.* 1987).

While there is evidence for cyclization of alphavirus and bunyavirus RNAs (Strauss and Strauss, 1983), the importance of cyclization in the replication of these viruses is unknown.

Studies of some RNA viruses have suggested that interactions between the 5'- and 3'-UTR sequences ('cross-talk' interaction) is essential for transcription, replication and packaging. For example, conserved sequences located at the 5'- and 3'-ends of influenza virus RNA segments (a negative - strand RNA virus) are complementary and act as *cis* regulatory elements in RNA synthesis (O'Neill and Palese, 1994). In the case of HCV, there is no evidence for base-pairing interactions between the 98-nucleotide element of 3'-UTR and upstream sequences (Blight and Rice, 1997) since addition of 5'- and 3'-UTR sequences or the full-length HCV genome did not modify the *in vitro* RNase cleavage pattern of the 98 nucleotide-3'-terminal region. However, the possibility that this type of interaction occurs *in vivo* cannot not be excluded.

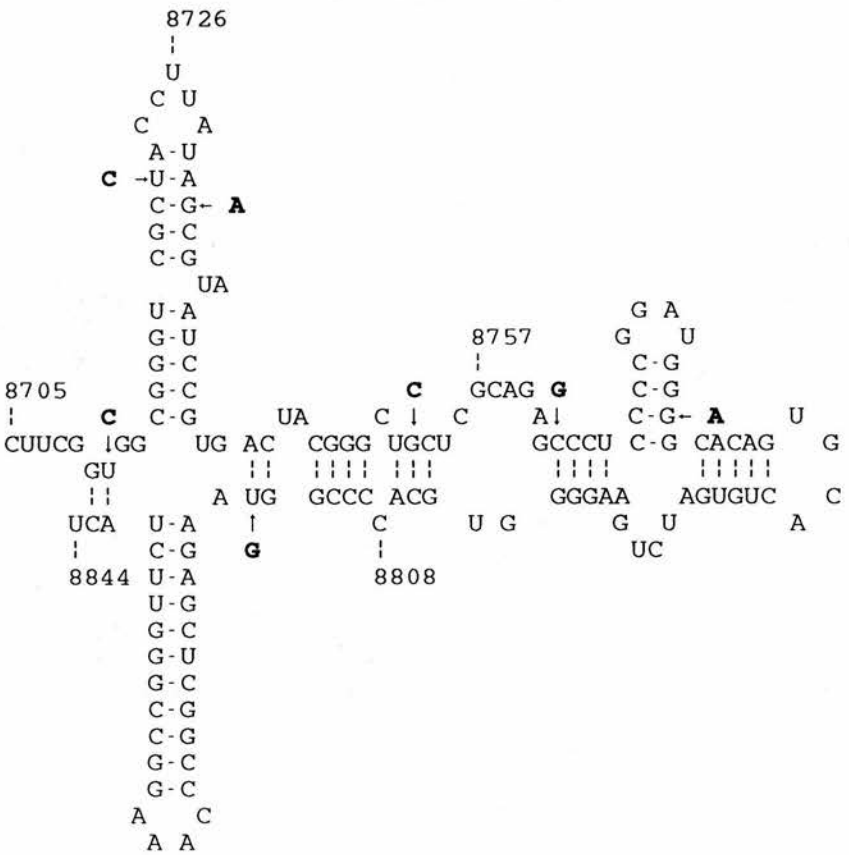
Secondary structure predictions combined with comparative sequence analysis of the 3'-UTR of the HGV/GBV-C genome have led to the construction

of a consensus model and the identification of a series of stable and conserved stem-loop structures (Figure 5.3). The conservation of these elements amongst thirty-nine isolates together with the presence of covariant substitutions suggest that these structures exist *in vivo* and are functionally important in viral RNA replication. At the same time, the free energy values of this secondary structure calculated for several HGV/GBV-C isolates of different phylogenetic groups indicated that the proposed structure is thermodynamically stable (Table 5.2). In addition, the RNAdraw program allowed the structure prediction and comparison of a number of 'scrambled sequences' for a specific HGV/GBV-C 3'-UTR. New files can be created with randomized portions of the sequence and used to assess whether changes in nucleotide order would have any effect on the free energies predicted by RNA folding. Sequence scrambling was carried out using different HGV/GBV-C isolates for both sense strand and reverse complement strand RNA (Table 5.2). The percentage reductions in free energy on sequence scrambling of the sense strand were significantly greater (mean reduction 19.37%) than those of the reverse complement strand (mean 13.74%; $p = 0.012$ by Mann-Whitney U test) (Table 5.2). The predicted secondary structure corresponding to the reverse complement strand, in which most of the hairpins are conserved amongst the isolates, is less stable than the conformation presented by the positive strand RNA (Table 5.2). Despite this, the models share a certain degree of structural similarity since hairpins **2** and **VI**, **3** and **Vb**, **5** and **IVb** are identical in size and position (Figure 5.5). A structural identity also exists between hairpins **11** and **Ib** as well between hairpin **IIIb** and the upper

region of hairpin **8** (Figures 5.3 and 5.5). However, the covariances present within hairpin **10** are not associated with the HGV/GBV-C groups to the same extent as those within hairpin **IIa** of the sense strand (Figure 5.6).

The consensus model proposed and analyzed in this chapter contradicts several aspects of a previously reported model based on the analysis of the last 140 nucleotides of the 3'-UTR from only three isolates (PNF2161, GT110, GT230) (Figure 5.7) (Okamoto *et al.* 1997). When thirty-nine HGV/GBV-C sequences described above were used to analyze the Okamoto *et al.* model, several substitutions were identified that disrupted the proposed base-pairings. For example, pairing of positions 8711 (U) and 8842 (A) is disrupted in all group 3, group 1 and unassigned sequences because of U→C substitution at position 8711 and in one group 2 sequence (Ed_34) because of a U→G substitution; another U-A pair between 8722 and 8730 is affected in three sequences (Ed_34, Png_29, Png_48) because of a U→C substitution. Other positions at which substitutions produce mismatches in regions predicted to be base-paired by Okamoto et al. are: C-C at positions 8753 and 8806 in four group 3 and two unassigned sequences; G-G at positions 8763 and 8802 in three group 3 and 2 unassigned sequences; A-G at positions 8744 and 8814 in one group 2 sequence (Figure 5.7). In addition, the Okamoto *et al.* model is not supported by any covariant substitutions and does not include the fragment upstream of position 8705 in which important structural elements could be identified and confirmed by either covariant substitutions or by highly conserved nucleotide sequences as presented in Figure 5.3.

Figure 5.7 Secondary structure model for the 3'-terminal 140 nucleotides of the 3'-UTR of HGV/GBV-C proposed by Okamoto et al. The substitutions responsible for disrupting the basepairing are shown in bold.



Computer-predicted folding patterns and RNase cleavage experiments have previously demonstrated the existence of a long stable hairpin structure (3'-LSH) within the distal part of the 3'-UTR of several different flaviviruses (Rice *et al.* 1985; Brinton *et al.* 1986; Proutski *et al.* 1997), some positive strand RNA plant viruses (Strauss and Strauss, 1983), HCV (Kolykhalov *et al.* 1996; Blight and Rice, 1997) and pestiviruses (Deng and Brock, 1993). Other studies have provided evidence for a specific interaction between the 3'-LSH of flaviviruses and host cellular proteins, components of the virus replication complex (Blackwell and Brinton, 1995) or have demonstrated a specific binding of cellular proteins to the 3'-terminal 98 nucleotides of the HCV RNA (Ito and Lai, 1997).

The analysis of 3'-UTR sequences presented here provides evidence for a similar, long stable hairpin (hairpin **VII**) at the 3'-end of the HGV/GBV-C genome that was present among all HGV/GBV-C isolates studied. This element has also been identified from analysis of a shorter fragment of the 3'-UTR, (Katayama *et al.* 1998) and is part of the structure proposed by Okamoto *et al.* (Figure 5.7). This stem-loop could be considered as a 3'-LSH and might be functionally important in RNA replication and/or in virus assembly. The existence of a long stable hairpin 3'LSH located at the 3'end of the genome RNA appears to be a common feature of the divergent members of *Flaviviridae* family (flaviviruses, pestiviruses, HCV and HGV/GBV-C viruses). At the same time, this structure might play a role in the recovery of infectious RNA transcripts from cloned cDNA since highly conserved structural elements and

correct terminal sequences are essential for virus replication (Boyer and Haenni, 1994; Yanagi *et al.* 1998).

The high conservation of the structure through covariance and free energy values amongst HGV/GBV-C isolates of different phylogenetic groups indicates that the possibility of the existence of such a structure *in vivo* that would be functionally important to the virus.

HGV/GBV-C, together with GBV-A and GBV-B, represent the GB viruses, recently discovered RNA viruses that appear to be members of the *Flaviviridae* family. Phylogenetic analysis of 5'-UTRs and structural genes showed that GBV-A and HGV/GBV-C share a common ancestor while GBV-B is more similar to HCV (Leary *et al.* 1996b). Comparison of nucleotide sequences of four GBV-A isolates available in GenBank suggests that their 3'-UTR sequences are incomplete, since they vary from 81 nucleotides (AF023245 isolate) to 197 nucleotides (HGU22303 isolate). As observed for HGV/GBV-C isolates, no poly(U) or poly(A) tails have been identified and instead GC-rich sequences are present. In contrast to the findings for HGV/GBV-C, folding predictions of the 3'-UTR sequence of different GBV-A isolates revealed very different secondary structure patterns, and these showed no significant similarity with the proposed structure of the 3'-UTR of HGV/GBV-C. These differences suggest that the 3'-terminal sequences of GBV-A isolates may be incomplete.

In conclusion, the analysis of the 3'-UTR of HGV/GBV-C RNA presented in this chapter has provided evidence for the existence of primary and secondary structure elements that may have potential biological functions in

important processes such as RNA replication, transcription and viral assembly. The existence of these structural elements remains to be confirmed by experimental analysis of RNA with single- and double-strand specific RNases. Their role in viral replication will have to be determined by functional assays.

CHAPTER 6

6. SUSCEPTIBILITY OF DIFFERENT TYPES OF CULTURED CELLS TO HGV/GBV-C AND HCV INFECTION

6.1 INTRODUCTION

The development of cell culture systems susceptible to infection and capable of supporting the viral replication is required in order to fully understand the mechanisms of the pathogenesis and persistence of HCV and HGV/GBV-C. Both viruses, distantly related to each other within *Flaviviridae* family, have single-stranded, positive-sense RNA genomes with a similar organization of gene products. By analogy with flaviviruses, the replication of the viral plus-strand RNAs is thought to occur via a minus-strand intermediate molecule. This minus-strand RNA intermediate serves as a template for synthesis of new genomic plus-strands which can be translated, used for further rounds of minus-strand synthesis or encapsidated into new virus particles.

Considerable efforts have been made to determine and characterize the cell tropism of HCV by identifying the virus localization at the cellular and subcellular levels (Bouffard *et al.* 1992; Zignego *et al.* 1992; Aria *et al.* 1993; Blight *et al.* 1993; Gabrielli *et al.* 1994; Lau *et al.* 1994; Gastaldi *et al.* 1995). Since over 60% of acute hepatitis C infection becomes chronic and often leads to liver cirrhosis and hepatocellular carcinoma (Gerber, 1995), the liver is regarded as the main target for replication of HCV *in vivo*.

Several assays such as *in situ* hybridization (ISH), *in situ* reverse

transcription PCR (IS-RT-PCR) and immunochemistry have been used for *in situ* detection of HCV genome and gene products in liver. For example, HCV RNA has been detected in the liver tissue samples of infected individuals by *in situ* hybridization using radiolabelled (Lamas *et al.* 1992; Haruna *et al.* 1993) or non-radioactive probes (Aria *et al.* 1993; Tanaka *et al.* 1993; Gastaldi *et al.* 1995). In addition, in some of these ISH studies, minus-strand HCV RNA, considered as a marker for virus replication, has also been detected (Lamas *et al.* 1992; Aria *et al.* 1993; Gastaldi *et al.* 1995).

The presence of HCV RNA in liver has also been demonstrated by *in situ* RT-PCR in which the PCR products can be detected by either direct incorporation of digoxigenin-labelled dUTP during amplification (Nuovo *et al.* 1993; Lau *et al.* 1994; Walker *et al.* 1998) or indirect detection based on the use of a labelled probe, as in ISH (Nuovo *et al.* 1993; Walker *et al.* 1998). At the same time, immunohistochemical techniques have been applied to detect and locate HCV antigens in liver tissue sections (Hiramatsu *et al.* 1992; Blight *et al.* 1993; Blight *et al.* 1994; Gonzalez-Peralta *et al.* 1994; Yap *et al.* 1994; Brody *et al.* 1998; Walker *et al.* 1998). Primary antibodies directed towards HCV viral antigens were applied on chemically fixed paraffin-embedded or frozen liver sections. Various HCV viral antigens such as core (Hiramatsu *et al.* 1992; Gonzalez-Peralta *et al.* 1994; Yap *et al.* 1994), E1 (Hiramatsu *et al.* 1992), E2 (Blight *et al.* 1994), NS3 (Hiramatsu *et al.* 1992; Blight *et al.* 1994), NS4 (Blight *et al.* 1993; Blight *et al.* 1994; Gonzalez-Peralta *et al.* 1994) and NS5 (Blight *et al.* 1994) were reported to be detected in different liver tissue samples using this

assay.

However, conflicting results and certain limitations have been identified based on all these techniques. Using ISH, various patterns of cellular and subcellular localization of HCV RNA within the liver have been reported. For example, HCV RNA was detected in the cytoplasm of hepatocytes (Lamas *et al.* 1992; Aria *et al.* 1993; Haruna *et al.* 1993; Tanaka *et al.* 1993; Lau *et al.* 1994; Gastaldi *et al.* 1995) and also in their nucleus (Blight *et al.* 1992; Lamas *et al.* 1992; Aria *et al.* 1993). In addition, positive signals indicating the presence of HCV RNA were detected in mononuclear cells (Blight *et al.* 1992; Lamas *et al.* 1992; Aria *et al.* 1993) and even in sinusoidal cells and bile duct epithelium (Aria *et al.* 1993). At the same time, the sensitivity of this technique is relatively low due to the possible degradation of the viral RNA during tissue preparation. Similar differences in the findings were revealed by IS-RT-PCR studies in which the HCV-specific positive signal appeared to be mainly cytoplasmatic (Lau *et al.* 1994) or mainly perinuclear (Nuovo *et al.* 1993) or intranuclear with either a diffuse or a peripheral pattern (Walker *et al.* 1998). In some cases, apart from hepatocytes, IS-RT-PCR also indicated the presence of HCV RNA in bile duct cells and portal lymphocytes (Walker *et al.* 1998), in Kupffer cells (Nuovo *et al.* 1993) and occasionally, in mononuclear cells (Lau *et al.* 1994). The diffusion of PCR products seemed to be a fairly common problem when the number of PCR cycles was more than fifteen (Lau *et al.* 1994), making the localization of HCV RNA difficult to assess. In addition, when indirect IS-RT-PCR was carried out, the signal was weak, affected by denaturation and stringent washing steps (Lau

et al. 1994).

In the case of immunohistochemical detection of HCV antigens in liver, the cellular and subcellular localization was consistent. Various HCV antigens-core, E1, NS3 (Hiramatsu *et al.* 1992), E2, NS3, NS4, NS5 (Blight *et al.* 1994), core, NS4 (Gonzalez-Peralta *et al.* 1994)-were always detected in the cytoplasm of hepatocytes. However, in only a few studies was the relationship between the hepatic expression of HCV antigens and pathobiological effect analyzed, revealing different conclusions. While no direct association was observed between the distribution of NS4-positive hepatocytes and areas of hepatocyte necrosis in liver biopsy specimens collected from anti-HCV positive patients (Blight *et al.* 1993), the HCV antigen expression in liver samples from chronic hepatitis C patients was correlated with disease activity and advanced disease stage (Hiramatsu *et al.* 1992).

The presence of HCV has also been investigated in extrahepatic sites such as peripheral blood mononuclear cells (PBMCs), bone marrow cells, spleen, pancreas, kidney and lymph nodes that could be capable of supporting the virus replication. For example, minus-strand RNA has been detected in PBMCs or subpopulations of PBMCs isolated from patients infected with HCV by *in situ* hybridization (Moldvay *et al.* 1994) and strand-specific RT-PCR (Bouffard *et al.* 1992; Zignego *et al.* 1992; Muller *et al.* 1993; Lerat *et al.* 1996) suggesting that HCV RNA might replicate in these cells.

Because the levels of HCV RNA found in infected individuals are relatively low, the RT-PCR is often considered the most reproducible and

sensitive technique to detect the presence of viral RNA in comparison to ISH. However, the strand specificity of RT-PCR using either sense or antisense primers for the detection of minus-strand RNA has been questioned. There is evidence that either false priming or self-priming of the plus-strand HCV RNA due to the hairpin loops of secondary structure present in the 5'-UTR leads to reverse transcription of the plus-strand RNA in the absence of antisense primers (Lanford *et al.* 1994; Mcguinness *et al.* 1994; Sangar and Carroll, 1998). In order to avoid these non-specific events, various modifications of strand-specific assays have been carried out. One such assay is tail or tagged PCR in which the primer used for cDNA synthesis contains an additional sequence at the 5'-end unrelated to any part of the virus genome (Chaves *et al.* 1994; Lanford *et al.* 1994). After reverse transcription, the tagged cDNA is amplified by PCR using only the tag sequence as one of the primers and an HCV-specific oligonucleotide as the opposing primer. Using this procedure, the strand-specificity is increased by preventing the amplification of cDNA products obtained by false priming of the incorrect RNA strand in conventional PCR. Alternatively, false priming can be avoided by using a thermoresistant RT (*rTth*) for cDNA synthesis at 70^0C (Lanford *et al.* 1994; Lanford *et al.* 1995). Since the strand specificity requires that no RT activity is present during PCR amplification, this is achieved by chelation of Mn^{2+} with EGTA while the DNA polymerase activity of the enzyme is activated by addition of Mg^{2+}. Another method used to overcome the nonspecific cDNA synthesis due to random priming by contaminating cellular nucleic acids is chemical modification of the 3'-end of RNA before the RT step

(Gunji *et al.* 1994). Under these conditions, RNA molecules with modified 3'-end (the 3'-terminal nucleotide is converted to a di-alcohol) cannot act as primers and the cDNA is specifically synthesized during RT only by addition of a specific primer. Another approach to avoid non-specificity in RT-PCR assays is to use primers that will amplify the core region of HCV genome (Lerat *et al.* 1996) or to use core primers in combination with the tag primer (Mellor *et al.* 1998). Primers specific to the HCV 5'-UTR has been widely used for amplification of HCV genomic strands because this region is the most highly conserved domain. However, most of the methods developed for the detection of negative strand HCV RNA based on amplification of 5'-UTR may generate false positive results due to self-priming events which the secondary structure of this region might be responsible for (Lanford *et al.* 1994; Tokita *et al.* 1994). The use of primers specific to conserved domains of HCV core region, in comparison with the tagged 5'-UTR primers, increased the overall specificity of the assay for the detection of negative strand RNA molecules (Lerat *et al.* 1996). In order to maximize the specificity of the detection method, primers for the core region that were highly conserved between published variants of HCV were designed and a tag sequence was added at the 5'-end of the outer primers (Mellor *et al.* 1998). The use of tagged core primers together with highly stringent amplification conditions (pre-denaturation of the template RNA at 70^0C for 10 min prior to the addition of AMV reverse transcriptase at 42^0C, heat-inactivation of the enzyme at 98^0C for 10 min after RT reaction followed by hot-start PCR) reduced further non-specific priming. Analysis of samples from HCV-

infected individuals using these improved strand-specific RT-PCR techniques suggests that HCV does not replicate in PMBCs, or only replicates at very low levels (Gunji *et al.* 1994; Lanford *et al.* 1995; Lerat *et al.* 1996; Laskus *et al.* 1997; Mellor *et al.* 1998). For example, no minus-strand HCV RNA was detected in PBMCs in any sample despite the presence of abundant plus-strand RNA using strand-specific *Tth*-based RT-PCR (Lanford *et al.* 1995; Laskus *et al.* 1997). Similarly, using RT-PCR method combined with chemical RNA modification at the 3'-end, minus-strand HCV RNA was detected in PBMCs in only one case (10%) while plus-strand RNA in eight cases (80%) (Gunji *et al.* 1994). In ten chronically HCV-infected patients, the highly stringent strand-specific PCR assay based on tagged core primers detected plus-strand HCV RNA in different PBMC subsets (mainly in B lymphocytes) and minus-strand RNA in a single subset (dendritic cells) of one of the samples analyzed (Mellor *et al.* 1998).

The presence of HCV in bone marrow cells has been also investigated using RT-PCR based methods (Gabrielli *et al.* 1994a; Galli *et al.* 1995a; Lanford *et al.* 1995a; Lerat *et al.* 1996a; Laskus *et al.* 1998a), but the findings are controversial. Although HCV replicative intermediates have been identified in bone marrow cells from patients with cryoglobulinemia (Gabrielli *et al.* 1994; Galli *et al.* 1995), no HCV negative strand RNA was detected in a bone marrow biopsy from a chronically infected chimpanzee (Lanford *et al.* 1995), in fresh bone marrow cells from chronic hepatitis C patients (Lerat *et al.* 1996) or in bone marrow collected postmortem from HCV-infected patients who died of acquired immunodeficiency syndrome (AIDS)-related complications (Laskus *et*

al. 1998a).

In the case of HGV/GBV-C, the degree of controversy is even higher. Initially, an association between HGV/GBV-C infection and acute liver disease in humans was suggested since viral RNA was detected in sera from patients with acute, post-transfusion nonA-E hepatitis (Linnen *et al.* 1996) and in 50% of patients with fulminant hepatitis from Germany (Heringlake *et al.* 1996). The same conclusion was reached in another study in which sera from 35% of Italian patients with acute nonA-E hepatitis and 39% with chronic nonA-E hepatitis were HGV/GBV-C positive (Fiordalisi *et al.* 1996). However, the association between HGV/GBV-C viraemia and hepatitis has become more controversial since in hepatitis patients infected only with HGV/GBV-C, neither the presence or titre of viral RNA is correlated with biochemical evidence of liver disease (Alter *et al.* 1997). To investigate whether HGV/GBV-C replicates in the liver of infected patients, the presence of HGV/GBV-C RNA was assessed and quantified by strand-specific PCR techniques involving either cDNA synthesis at a high temperature with thermostable enzyme *Tth* (Laskus *et al.* 1997b; Laskus *et al.* 1998b), use of specific primers for 5'-UTR (Kudo *et al.* 1997), chemical modification of the 3'-end of RNA (Madejon *et al.* 1997; Saito *et al.* 1997; Kanda *et al.* 1998) or tagged primers specific for the NS3 region (Mellor *et al.* 1998). Detection of negative strand RNA would be considered as evidence of viral replication. However, in liver samples from HCV and HGV/GBV-C-coinfected patients, the plus strand of HGV/GBV-C RNA was only detected at very low levels compared to those of HCV RNA (Kudo *et al.* 1997) while the

minus strand was not found (Laskus *et al.* 1997; Kanda *et al.* 1998; Laskus *et al.* 1998; Mellor *et al.* 1998). These findings suggest that HGV/GBV-C, unlike HCV, is not hepatotropic, implying that its cell tropism might be different to that of HCV. In contrast, other reports indicated the presence of negative strand HGV/GBV-C RNA in liver biopsy samples using chemical modification method (Madejon *et al.* 1997; Saito *et al.* 1997). Since there is a limitation of this assay for detecting strand specificity when large amounts of HGV/GBV-C are present in the sample (Kanda *et al.* 1998), these results might not be accurate.

Since the discovery of HCV in 1989, many attempts have been made to develop an efficient system for HCV replication in cell culture either by infection or transfection.

In the infection experiments, serum or plasma collected from HCV-infected individuals has been used to infect cultured cells such as MOLT-4 human T-cell line (Shimizu *et al.* 1992), Daudi cells (B-cell line) (Nakajima *et al.* 1996), chimpanzee hepatocytes (Lanford *et al.* 1994), PBMCs from healthy donors (Cribier *et al.* 1995), MT-2 cells (human T-cell leukaemia virus type I-infected cell line) (Kato *et al.* 1995; Mizutani *et al.* 1996), HepG2 and HuH7 cells (human hepatoma cell lines) (Seipp *et al.* 1997), HFH cells (human foetal hepatocytes) (Iacovacci *et al.* 1997) and normal human hepatocytes in primary culture (Fournier *et al.* 1998). Although in some cases, long-term infection of up to 198 days (Mizutani *et al.* 1996) or more than 1 year (Mizutani *et al.* 1996) has been achieved, in general, HCV replication in cultured cells is transient or intermittent. Various reports have described different procedures for the

replication of HCV *in vitro* including: shift of temperature for culture incubation from 37^0C to 32^0C after virus inoculation (Mizutani *et al.* 1996), use of a large number of serum samples as inocula (Fournier *et al.* 1998), stimulation of LDL receptor expression and use of highly supplemented culture medium (Seipp *et al.* 1997). In most cases, detection of minus-strand HCV RNA by strand-specific PCR (Lanford *et al.* 1994; Cribier *et al.* 1995; Fournier *et al.* 1998) or nested-PCR followed by hybridization (Shimizu *et al.* 1992; Seipp *et al.* 1997) was considered as evidence of viral replication. The expression of virus-encoded proteins was also tested by examining the cultured cells for putative HCV core and NS4 antigens using indirect immunofluorescence method with mouse monoclonal anti-HCV antibodies (Shimizu *et al.* 1992). However, these investigations have been hampered by the limited specificity and sensitivity of methods used to detect plus- and minus-strand HCV RNA, by the low titres of RNA present, and by variation in infectivity between serum samples.

The alternative approach used to establish an *in vitro* permissable HCV culture system is by transfection of infectious RNA transcripts into HuH7 cells (Yoo *et al.* 1995) and HepG2 cells (Dash *et al.* 1997). In both reports, not only the full-length HCV RNA transcript but also HCV transcripts without the highly conserved 3'-end 98 nucleotide-sequence were found to be infectious. The interpretation of these results must be done with care since the last 46 nucleotides of this conserved region of HCV RNA were predicted to form a stable stem-loop structure considered to be critical for viral replication (Kolykhalov *et al.* 1996; Blight and Rice, 1997) and transcripts of a chimeric full-

length cDNA clone of HCV proved to be infectious *in vivo* when the infectivity was tested by intrahepatic transfection of a chimpanzee with transcribed RNA (Yanagi *et al.* 1998).

In contrast, efforts to develop reliable cell culture systems that are permissive to *in vitro* replication of HGV/GBV-C are at a very early stage. Two cell lines, MT-2C (a human T-cell leukaemia virus type I-infected cloned T cell line) and PH5CH (a non-neoplastic human hepatocyte line), have been reported as susceptible to HGV/GBV-C infection and capable of supporting the viral replication when the temperature was shifted from 37^0C to 32^0C after the virus inoculation (Ikeda *et al.* 1997).

The aims of the study described in this chapter were to examine the susceptibility of three types of cultured cells (PBMCs, U937 and HepG2 cells) to HCV and HGV/GBV-C infection and to identify the potential factors that could mediate the infection process in an attempt to develop reliable cell culture systems. Different approaches were used to infect the cells and a modified nested PCR-based assay using tagged primers in the core region of HCV and NS3 region of HGV/GBV-C that allows specific strand detection of single-strand RNA (Mellor *et al.* 1998) was performed.

PBMCs freshly isolated from the blood of healthy donors and two immortalized cell lines-the human monocytic leukaemia cell line U937 and the human hepatoma cell line HepG2-were tested for their susceptibility to HCV and HGV/GBV-C infection .

PMBCs were separated from heparinized blood using a Ficoll density gradient and resuspended in RPMI 1640 medium containing 10% FCS, antibiotics (10^4IU/ml), L-Glutamine (100mM) and IL-2 (10^4IU/ml) (section 2.5.1). In order to stimulate cell division, the cells were incubated with PHA (5μg/ml) for 48h before virus inoculation.

U937 cells grow as a single cell suspension and they have been shown to have characteristics of an undifferentiated monocytic cell (Sundstrom and Nilsson, 1976). These cells are variable in shape (round and polygonal) and can be induced to differentiate into macrophages by a variety of agents such as dimethylsulfoxide (DMSO) (Kay *et al.* 1983), vitamin D (Rigby *et al.* 1984) or phorbol ester (Lewis *et al.* 1987). In this study, the differentiation of U937 cells was achieved by the addition of phorbol 12-myristate, 13-acetate (PMA) to cultured cells before virus inoculation (section 2.5.1). Differentiated cells become adherent to tissue culture plastic and monolayers were incubated with HCV or HGV/GBV-C positive sera in separate flasks.

The human hepatoma cell line HepG2 was maintained in RPMI 1640 medium supplemented with 10% FCS and antibiotics (10^4IU/ml) as monolayers

in culture flasks at 37^{O}C in a 5% CO_2 atmosphere (section 2.5.1). HepG2 cells bind, take up and degrade low-density lipoproteins (LDL) through specific LDL receptors similar to those demonstrated on extrahepatic tissue cells (Havekes *et al.* 1983). Since HCV and HGV/GBV-C were found to be associated with low density lipoproteins in human plasma (Chapter 4), a putative LDL-mediated entry of the viruses into the cells was examined. For this purpose, the cells were grown in FCS-free Ham's F12 medium with supplements and lovastatin before virus inoculation (section 2.5.1). Lovastatin is an inhibitor of one of the enzymes involved in the cholesterol biosynthesis pathway HMG-CoA (3-hydroxy-3-methylglutaryl-coenzyme A reductase), with the result that the LDL-receptor on the cell surface is upregulated.

HCV and HGV/GBV-C positive serum samples with high viral RNA titres together with the negative control of a HCV and HGV/GBV-C negative serum were used for *in vitro* infection of the cells (Table 6.1). All experiments were carried out in duplicate.

In order to identify parameters that might increase the permissiveness of cells to viral infection, a number of different culture conditions were analysed. For example, different volumes of the inoculum were used during the *in vitro* infection step: 10 μl for PBMCs, 25 μl or 100 μl for U937 cells and 100 μl for HepG2 cells. The incubation period of cells with the viral inoculum was either 4h (PBMCs, U937), 24h (PBMCs, U937) or 90min (HepG2) at 37^{0}C in a 5% CO_2 atmosphere. Inoculum was then removed (day 0) and cells were washed three times with PBS and fresh medium was added. In all the experiments, at

Table 6.1 Serum samples used as inocula for *in vitro* infection of cultured cells

Sample	HCV RNA (copies/ml serum)	HGV/GBV-C RNA	Target cells
S1 (haemophiliac)	1.0×10^6	-ve	PBMC
S2 (haemophiliac)	-ve	1.0×10^6	PBMC
S3 (haemophiliac)	1.0×10^5	-ve	U937
S4 (blood donor)	-ve	5.0×10^8	U937
S5 (haemophiliac)	1.0×10^4	-ve	U937
S6 (haemophiliac)	1.0×10^6	-ve	HepG2
S7 (haemophiliac)	1.0×10^7	-ve	HepG2
S8 (blood donor)	-ve	1.5×10^7	HepG2
S9 (blood donor)	-ve	3.0×10^6	HepG2
S10 (blood donor)	-ve	-ve	all

-ve = PCR-negative
HCV-positive samples were quantified by J. Ellender (University of Edinburgh).
HGV/GBV-C-positive samples were quantified by F. Davidson (BTS-Edinburgh).

various times during the culture period (days post-inoculation), 1/10 of the inoculated culture was harvested and separated by centrifugation at 1000rpm for 5min. The supernatants and cell pellets were kept at -40^{0}C for RNA extraction (sections 2.2.1 and 2.2.2). HCV plus-strand RNA was detected by RT-PCR using primers specific to the 5'-UTR and by strand-specific PCR using tagged primers specific to the core region, while the presence of minus-strand was examined using tagged primers specific to the core region (Table 2.1). Similarly, the presence of HGV/GBV-C plus-strand RNA in the supernatants and cells of HGV/GBV-C infected cultures was investigated by RT-PCR using primers specific to the HGV/GBV-C 5'-UTR and by strand-specific PCR using tagged primers specific to the NS3 region while the minus-strand RNA was detected using tagged primers specific to the HGV/GBV-C NS3 region (Table 2.1). Reaction conditions for conventional RT-PCR and for strand-specific RT-PCR were as described in sections 2.2.4 and 2.2.5. Negative controls using HCV and HGV/GBV-C negative serum were included in each experiment and supernatants and cells were collected in parallel from cultures inoculated with this serum.

6.3 RESULTS

6.3.1 *IN VITRO* INFECTION OF PBMCs

PHA-stimulated PBMCs (1 x 10^6 cells/ml) were inoculated with three separate serum samples for 4h or 24 hr at 37^0C. In both cases, cell growth was not affected by inoculation. Cultures were harvested on days 1, 3 and 8 post-inoculation (p.i.) for the 4h-incubation experiment (Table 6.2 A) and on days 1, 2, 7 and 10 p.i. for the 24h-incubation experiment (Table 6.2 B). In both experiments, plus-strands of HCV RNA and HGV/GBV-C RNA were detected in supernatants collected after inoculation (Table 6.2 A, B). Plus-strand HCV RNA was also detected in cells harvested at day 1 p.i. for the 4h-incubation experiment (Table 6.2 A) and in cells at day 7 p.i. for the 24h-incubation experiment (Table 6.2 B). Minus-strands of HCV or HGV/GBV-C RNA were not detected in either the supernatants or cells of any of these cultures. HCV and HGV/GBV-C RNA were uniformly absent in the negative control cultures. Similar results were obtained when the 4h-incubation experiment was repeated when PHA-stimulated PBMCs (1 x 10^6 cells/ml) were inoculated with 2 different HCV and HGV/GBV-C positive serum samples (data not shown).

6.3.2 *IN VITRO* INFECTION OF U937 CELLS

The susceptibility of U937 cells to infection by HCV and HGV/GBV-C was investigated in two separate experiments using undifferentiated U937 cells

Table 6.2 Detection of HCV RNA and HGV/GBV-C RNA in PMBC culture. 1x 10^6 cells/ml inoculated with 10μl serum S1 (HCV) or S2 (HGV/GBV-C).

A. 4h-incubation experiment

	Supernatants				Cells			
	HCV RNA		HGV RNA		HCV RNA		HGV RNA	
	plus	minus	plus	minus	plus	minus	plus	minus
days p.i.								
0	+	-	+	-	nd	nd	nd	nd
1	-	-	-	-	+	-	-	-
3	-	-	-	-	-	-	-	-
8	-	-	-	-	-	-	-	-

B. 24h-incubation experiment

	Supernatants				Cells			
	HCV RNA		HGV RNA		HCV RNA		HGV RNA	
	plus	minus	plus	minus	plus	minus	plus	minus
days p.i.								
1	+	-	+	-	-	-	-	-
2	-	-	-	-	-	-	-	-
7	-	-	-	-	+	-	-	-
10	-	-	-	-	-	-	-	-

nd = not done
plus = plus-strand RNA
minus.=minus-strand RNA
p.i.=post-inoculation

(cell suspension) and differentiated (PMA-treated) U937 cells (monolayers). Since the differentiation of U937 cells increased permissiveness for haemorrhagic fever virus infection and production (Lewis *et al.* 1987), monolayers were prepared and used, in parallel with cell suspension, to determine any possible effect of cell differentiation on *in vitro* HCV or HGV/GBV-C infection. In both cases, cells were inoculated separately with either HCV or HGV/GBV-C positive serum samples.

Plus-strand HGV/GBV-C RNA could only be detected in PMA-treated U937 cells at day 1 p.i. and was absent from other cell or supernatant samples up to 4 days p.i. (Table 6.3 A). For the untreated non-adherent cells, HGV/GBV-C plus-strand RNA was only detected in the supernatant 1 day p.i. but not in other supernatant or cell samples harvested up to 4 days p.i. (Table 6.3 B). Plus-strand HCV RNA was only detected in cells and/or supernatants collected at days 2, 3 and 4 p.i. from the untreated cells, but no minus-strand RNA was detected up to 4 days p.i. (Table 6.3 B).

These findings were further investigated by maintaining a suspension of U937 cells in culture for 18 days p.i. The failure to detect minus-strand RNA in the cells during short-term culture (Table 6.3 B) suggested that the entry of the virus particles into the cell might require a longer exposure time (25 μl or 100 μl inoculum incubated for 24h at 37^{O}C, in parallel experiments). Plus-strand HCV RNA was detected at 15 days p.i. in cells inoculated with either of HCV-positive inocula (S3 or S5) and also at day 18 p.i. for inoculum S5 (Table 6.4 B). However, no minus-strand RNA was found in either the supernatants or cells

Table 6.3 Detection of HCV RNA and HGV/GBV-C RNA in U937 cells

A. PMA-treated U937 cells incubated with 100μl inoculum (HCV-S3 or HGV-S4) for 4h at 37^{0}C

	Supernatants				Cells			
	HCV RNA		HGV RNA		HCV RNA		HGV RNA	
	plus	minus	plus	minus	plus	minus	plus	minus
days p.i.								
1	-	-	-	-	-	-	+	-
2	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-
4	-	-	-	-	-	-	-	-

B. U937 cell suspension (2 x 10^5 cells/ml) incubated with 25μl inoculum (HCV-S3 or HGV-S4) for 4h at 37^{0}C.

	Supernatants				Cells			
	HCV RNA		HGV RNA		HCV RNA		HGV RNA	
	plus	minus	plus	minus	plus	minus	plus	minus
days p.i.								
1	nd	nd	+	-	nd	nd	-	-
2	+	-	-	-	+	-	-	-
3	+	-	-	-	-	-	-	-
4	+	-	nd	nd	+	-	-	-

nd= not done
plus = plus-strand RNA
minus = minus-strand RNA

Table 6.4 Detection of HCV RNA and HGV/GBV-C RNA in U937 cell suspension. 25μl or 100μl inoculum incubated with 2 x 10^5cells/ml for 24h at 37^0C.

A.

inoculum	S3		S5		S4	
	Supernatants					
	HCV RNA		HCV RNA		HGV RNA	
	plus	minus	plus	minus	plus	minus
days p.i.						
0	nd	nd	nd	nd	+	-
3	-	-	-	-	-	-
9	-	-	-	-	-	-
15	-	-	-	-	-	-
18	-	-	-	-	-	-

B.

	Cells					
	HCV RNA		HCV RNA		HGV RNA	
	plus	minus	plus	minus	plus	minus
days p.i.						
3	-	-	-	-	-	-
9	-	-	-	-	-	-
15	+	-	+	-	-	-
18	-	-	+	-	-	-

from these cultures (Tables 6.4 A and B).

For U937 cell suspensions exposed to HGV/GBV-C inoculum, only plus-strand RNA was detected in the supernatants immediately after inoculation (day 0) but not in later samples or in the cells (Table 6.4 A).

No HCV or HGV/GBV-C RNA was detected in either the supernatants or cells harvested from U937 monolayers or cell suspensions inoculated with the negative control serum.

6.3.3 *IN VITRO* INFECTION OF HepG2 CELLS

Prior to virus inoculation, HepG2 monolayers were harvested by trypsinisation and the cells washed and resuspended (1 x 10^6 cells/ml) in FCS-free Ham's F12 medium with supplements and treated with lovastatin (final concentration 10 μM) in order to stimulate the expression of the LDL receptor on the cell surface (section 2.5.1). The HepG2 monolayers were then inoculated with 100 μl of HCV-(S6, S7) or HGV/GBV-C-(S8, S9) positive serum samples for 90 min at 37^0C. The inoculated monolayers were grown in serum-free highly supplemented Ham's F12 medium and harvested on days 0, 3, 7, 14 and 21 p.i.

All samples, at all times, were negative for HCV or HGV/GBV-C minus-strand RNA. The presence of plus-strands of HCV RNA and HGV/GBV-C RNA could only be detected in the supernatants collected immediately after inoculation (day 0) and day 3 p.i. (Table 6.5 A). Cells harvested during the culture period remained PCR negative for the presence of HCV RNA and

Table 6.5 Detection of HCV RNA and HGV/GBV-C RNA in HepG2 cells. Lovastatin-treated HepG2 cell monolayers incubated with 100μl inoculum for 90min at 37^{0}C.

A.

inoculum S6		S7		S8		S9	
			Supernatants				
HCV RNA plus	HCV RNA minus	HCV RNA plus	HCV RNA minus	HGV RNA plus	HGV RNA minus	HGV RNA plus	HGV RNA minus
days p.i.							
0	+	-	+	-	+	-	+	-
3	+	-	-	-	+	-	-	-
7	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-

B.

			Cells				
HCV RNA plus	HCV RNA minus	HCV RNA plus	HCV RNA minus	HGV RNA plus	HGV RNA minus	HGV RNA plus	HGV RNA minus
days p.i.							
0	-	-	-	-	-	-	-	-
3	-	-	-	-	-	-	-	-
7	-	-	-	-	-	-	-	-
14	-	-	-	-	-	-	-	-
21	-	-	-	-	-	-	-	-

HGV/GBV-C RNA, respectively (Table 6.5 B).

6.4 DISCUSSION

Establishment of easily maintained culture systems for *in vitro* infections with HCV or HGV/GBV-C is critical for investigation of virus-cell interaction and other important aspects related to replication and pathogenesis of these viruses. In this study, three types of cultured cells were inoculated with serum samples from HCV- or HGV/GBV-C infected individuals in order to investigate their susceptibility to infection. The sera selected as inocula contained high viral RNA titres (Table 6.1) and various culture conditions were applied before or during the virus inoculation to increase the putative permissiveness of these cells to virus infection. However, different volumes of inoculum, variable incubation times, induction of cell differentiation before virus inoculation or stimulation of the LDL receptor expression had no significant effect in supporting the *in vitro* replication of either HCV or HGV/GBV-C. The presence of HCV RNA or HGV/GBV-C RNA in supernatants or cells after inoculation (day 0) and during the first days of culture in PBMCs (Table 6.2), in U937 (Tables 6.3, 6.4) and HepG2 cells (Table 6.5) might be ascribed to the inoculum or non-specific attachment/adhesion of virus particles to the cell surface.

The lack of evidence for the *in vitro* replication of HCV and HGV/GBV-C in any of the cultured cells used in this study could be analyzed considering the following aspects:

a) Type of inoculum used.

Compared to hepatitis B virus, the levels of HCV and HGV/GBV-C RNA in infected individuals are relatively lower and their routine detection normally relies on nested RT-PCR. In this study, serum samples with high levels of viral RNA were used as inocula (Table 6.1), in order to maximise the chance of the cells becoming infected *in vitro*. However, although PCR is a sensitive detection method, it cannot discriminate between infectious and inactivated virus particles, and the infectivity of the inoculum will not necessarily correlate with the level of viral RNA. This possibility is reinforced by experiments in which a number of sera from chronically HCV-infected individuals were screened for *in vitro* infection of normal chimpanzee hepatocytes (Lanford *et al.* 1994) or primary cultures of adult human hepatocytes (Fournier *et al.* 1998). In the latter study, evidence of infection of the cells with HCV was obtained using strand-specific r*Tth* RT-PCR for only 10 out of the 33 sera, and three of the non-infectious inocula had high RNA titres (1.12×10^8-1.41×10^8 copies/ml). To avoid this problem, multiple serum samples (5 HCV- and 4 HGV/GBV-C positive) were used in this study but no evidence for replication was obtained.

b) Sensitivity and specificity of RT-PCR assay.

The strand-specific RT-PCR assay based on detection of the negative strand intermediate molecules of HCV and HGV/GBV-C has been used in most of the studies that investigated the viral replication in the infected cells *in vivo* and *in vitro* (Lanford *et al.* 1994; Lanford *et al.* 1995; Lerat *et al.* 1996; Kudo *et al.* 1997; Laskus *et al.* 1997; Saito *et al.* 1997; Mellor *et al.* 1998). Potential causes

of non-specific amplification of both positive and negative strands of viral RNA that could lead to false positive results have been considered and examined (Gunji *et al.* 1994; Kudo *et al.* 1997; Mellor *et al.* 1998; Sangar and Carroll, 1998).

The specificity and sensitivity of strand-specific RT-PCR assays used in this study have been assessed and validated in a previous report which investigated whether sub-populations of PBMCs could be considered as possible extra-hepatic sites for *in vivo* replication of HCV and HGV/GBV-C (Mellor *et al.* 1998). To maximise the two essential characteristics of the detection, amplification conditions were highly stringent: the template RNA was pre-denatured at 70^0C for 10min prior to the addition of reverse transcriptase, and after the cDNA synthesis the samples were heated at 98^0C for 2 min to heat-inactivate the RT enzyme, 'hot-start' PCR with an elevated annealing temperature of 55^OC for 30 cycles was carried out (Mellor *et al.* 1998) (section 2.2.4). At the same time, to maximise the specificity, PCR was performed using newly designed tagged primers specific for the core region of HCV or NS3 region of HGV/GBV-C (Table 2.1). This strand-specific detection system, validated using synthetic RNA transcripts for each strand of both viruses, indicated that the non-specific priming that could be attributed to residual plasmid DNA, was significantly reduced (no positive signals detected in the titration of synthetic HCV transcripts serially diluted up to $1:10^{10}$ or of HGV/GBV-C transcripts up to $1:10^7$) (Mellor *et al.* 1998).

Using this strand-specific PCR assay, no negative strand of HCV RNA

or HGV/GBV-C RNA could be detected in the supernatants or cells harvested at different times after inoculation of PMBCs, U937 or HepG2 cells. Although plus-strand HCV RNA was detected at days 15 and 18 p.i. in the cell pellets from U937 cell suspensions (Table 6.4 B), the failure to detect minus-strand or plus-strand of HCV RNA released as new virions in supernatants at the same time suggests that this does not reflect viral replication. However, the possibility of a low level of replication cannot be totally excluded since no RNA quantification of PCR-positive samples was carried out. As previously suggested, low-level replication or replication limited to a small number of cells would remain undetected because minus-strand HCV RNA is generally detected, in cells supporting replication, at 100- to 1000-fold-lower levels than plus-strand RNA (Lanford *et al.* 1995).

c) Cell tropism of HCV and HGV/GBV-C.

The question of whether PBMCs can be infected by HCV *in vivo* or *in vitro* has been a controversial issue mainly due to the methodology employed to detect minus-strand RNA molecule, the replicative intermediate (Zignego *et al.* 1992; Muller *et al.* 1993; Cribier *et al.* 1995; Lanford *et al.* 1995; Lerat *et al.* 1996; Laskus *et al.* 1997; Mellor *et al.* 1998). However, most of the reports that used strand-specific PCR-based assays of sufficient sensitivity and specificity conclude that if HCV replication occurs in PBMCs, it is only at a very low level.

Following the observation that PBMCs are permissive for HCV replication *in vitro* (Cribier *et al.* 1995), their susceptibility to HCV and HGV/GBV-C infection was tested in this study by incubating PMBCs isolated

from healthy donors with HCV- or HGV/GBV-C-positive sera. The lack of evidence for *in vitro* replication of HCV in PBMCs (Table 6.2) is consistent with previously published findings suggesting that the rate of HCV replication in PBMCs inoculated *in vitro* is low and that virus remains primarily cell-associated (Cribier *et al.* 1995). Similarly, in the case of HGV/GBV-C, PBMCs could not be infected *in vitro* (Table 6.2) in agreement with the lack of evidence for HGV/GBV-C replication *in vivo* in PBMCs collected from HGV/GBV-C infected individuals (Laskus *et al.* 1998; Mellor *et al.* 1998; Radkowski *et al.* 1998).

The second type of cultured cells used in this study (U937 cells) is a continuous cell line that exhibits morphologic and functional characteristics of undifferentiated monocytic cells (Sundstrom and Nilsson, 1976). These cells have proved useful for studies of viral infection with Herpes simplex virus type 1 (Linnavuori and Hovi, 1981), measles virus (Brandriss *et al.* 1982), and haemorrhagic fever virus (Lewis *et al.* 1987). In case of HCV, the presence of viral RNA and core antigen was detected in a monocyte/macrophage population of PBMCs from infected patients (Bouffard *et al.* 1992).

Previously, various cell lines were tested for their susceptibility to HCV infection, and U937 cells were found to be relatively susceptible because positive strand RNA was detected in cells 7 days p.i. (Kato *et al.* 1995). In reality, the detection of positive strand HCV RNA, especially during the first days of culture period, could simply indicate the adsorption of the viral particles to the cell membrane and cannot be considered as evidence for viral replication. Consistent

with this observation, under the experimental conditions tested, U937 cells were not susceptible to HCV or to HGV/GBV-C infection (Table 6.4). Replication was also not observed when U937 cells were induced to differentiate to macrophage-like cells by phorbol esters (PMA) (Table 6.3 A) although the degree of cellular differentiation seems to be a determining factor of the susceptibility of these cells for infection by viruses (Lewis *et al.* 1987). Differentiation of early monocytic precursors to more mature monocytes that resemble macrophages using phorbol esters was previously shown to be a membrane-mediated process (Cooper *et al.* 1982). The sequence of events by which the changes occur is not well understood. However, the inducer appears to be bound to a specific surface receptor on the cells and it is not internalized (Cooper *et al.* 1982). For HCV or HGV/GBV-C, PMA-induced alterations of U937 cells did not facilitate the viral entry, an essential step in cell-virus interaction.

HepG2 cells are well differentiated hepatoblastoma cells that display certain functions similar to human hepatocytes such as synthesis of a number of plasma proteins and apopoliproteins and the presence of high-affinity LDL receptors (Havekes *et al.* 1983). The hypothesis that the LDL receptor present on the surface of hepatocytes facilitates the entry of HCV into cells is supported by the observation that positive strand HCV RNA could be detected for 50 days p.i. when LDL receptor expression was stimulated by lovastatin in HCV-infected HuH7 cells (human hepatoma cell line) grown in FCS-free supplemented Ham's F12 medium (Seipp *et al.* 1997).

However, the HepG2 cells examined in this study did not show susceptibility to either HCV or HGV/GBV-C infection when LDL receptor expression was stimulated (Table 6.5).

The replication of HGV/GBV-C in cultured human cells (MT-2C and PH5CH) has been reported (Ikeda *et al.* 1997). These cell types have been previously found to be capable of supporting the replication of HCV (Mizutani *et al.* 1996). After inoculating the cells with serum containing high titres of both HCV RNA and HGV/GBV-C RNA, a shift of temperature from 37^0C to 32^0C seemed to be effective in supporting HGV/GBV-C as well as HCV replication. HCV RNA and HGV/GBV-C RNA could be detected in both cell lines more than 30 days p.i. by RT-nested PCR using primers from 5'-UTR of HCV RNA and HGV/GBV-C RNA and from 3'-UTR of HGV/GBV-C RNA. However, the presence of negative strand RNA of either HCV or HGV/GBV-C was not examined in this study. As previously suggested, standard RT-PCR is not strand-specific because of false priming of the incorrect strand or self-priming related to RNA secondary structures (Lanford *et al.* 1994). Since, generally, the presence of replicative intermediates must be established as evidence for replication (Cribier *et al.* 1995; Lanford *et al.* 1995; Laskus *et al.* 1997; Laskus *et al.* 1997; Seipp *et al.* 1997; Fournier *et al.* 1998; Mellor *et al.* 1998), strand-specific RNA detection should be performed before drawing a final conclusion about HCV or HGV/GBV-C replication in these cultured cells.

In conclusion, the lack of evidence for HGV/GBV-C replication in certain

cultured cells (PBMCs, U937 and HepG2 cells) under experimental conditions tested together with previous studies of *in vivo* replication (Laskus *et al.* 1997; Kanda *et al.* 1998; Laskus *et al.* 1998; Mellor *et al.* 1998; Radkowski *et al.* 1998) suggest that the identification of the *in vivo* sites of replication may be necessary before it is possible to establish a truly permissive cell system for the *in vitro* culture of HGV/GBV-C.

CHAPTER 7

7. FINAL DISCUSSION

Since the discovery and isolation of HGV/GBV-C in 1995 (Simons *et al.* 1995a; Linnen *et al.* 1996a), there has been considerable interest in the prevalence and clinical characteristics of HGV/GBV-C infection. The virus can cause acute and persistent infection in humans and has a worldwide distribution (Europe, Asia, South America, Africa), being present in both healthy and diseased population groups (Chapter 1). However, the pathogenicity of HGV/GBV-C is controversial since in most cases, infection appears to be asymptomatic with slight or no increase of serum ALT levels and the causative role of HGV/GBV-C in hepatitis is increasingly being questioned.

The aim of this study was to investigate and analyse the heterogeneity of HGV/GBV-C and to characterize the terminal regions of the viral genome. Comparative analysis of nucleotide sequence data revealed a very different pattern of HGV/GBV-C variability from that identified in HCV. For example, it is possible to classify HCV isolates into six major genotypes which can be divided further in subtypes, on the basis of sequence comparisons of complete genomes or of subgenomic regions (Simmonds *et al.* 1993a; Simmonds *et al.* 1994a). At the same time, it is known that infection with different HCV genotypes is associated with important clinical differences in serological responses, pathogenicity, and the response to interferon treatment (Dusheiko *et al.* 1994; Simmonds, 1997). The extensive phylogenetic analysis carried out in this study, in which a large number of HGV/GBV-C full-length and 5'-UTR

sequences have been compared has demonstrated the presence of four phylogenetic groups of isolates which are well correlated with their geographical origin (Chapter 3). Although the diversity between HGV/GBV-C variants is limited in comparison to that between HCV types, phylogenetic groupings of HGV/GBV-C can be reproduced by analysis of the 5'-UTR and of sub-fragments.

This information about virus heterogeneity and the geographical distribution of HGV/GBV-C groups is relevant to discussions about the origin of this virus. For example, in the case of HCV, the existence of numerous subtypes in a single geographical area is consistent with the long-term presence of HCV in these populations (Mellor *et al.* 1995; Smith *et al.* 1997). The rate of HCV sequence change measured amongst a cohort of individuals infected with HCV following administration of anti-D immunoglobulin was calculated at 4.1 x 10^{-4} per site per year for NS5B and 7.4 x 10^{-4} per site per year for E1, while the divergence of HCV types was estimated to occur more than 500 years ago (Smith *et al.* 1997). A similarly high rate of sequence change of 3.9 x 10^{-4} per site per year was estimated for the whole genome of HGV/GBV-C over 8.4 years (Nakao *et al.* 1997). Since HGV/GBV-C variants are more similar to each other than HCV types, this rate of substitution would imply that HGV/GBV-C variants have emerged within a few hundred years.

Another approach to elucidating the origin of HGV/GBV-C is based on construction and analysis of phylogenetic trees of the NS3 and NS5A regions from different HGV/GBV-C isolates (Tanaka *et al.* 1998). Using the New World

monkey isolate GBV-A as an outgroup, a distinct subgroup of group 1 of isolates from African pygmies and containing a duplicated region within the NS5A gene, appeared to be ancestral, interpreted by the authors as indicating an African origin of GBV-C. However, more recent analysis based upon complete genome sequences and in which the chimpanzee isolate HGV/GBV-C_{tro} was used as an outgroup have failed to confirm that group 1 isolates are ancestral (D.B.Smith, personal communication). The HGV/GBV-C_{tro} is a novel flavivirus recently detected in non-captive chimpanzees from Central and West Africa (Adams *et al.* 1998). Sequence comparisons of 5'-UTR, NS3 and NS5 regions showed that this virus is distinct from both HGV/GBV-C and GBV-A, but more closely related to human viruses. Distinct variants of this virus, recovered from different subspecies of chimpanzees are more diverse than variants of HGV/GBV-C found in humans. For example, 19.3% nucleotide and 9.5% amino acid sequence diversity in the NS5 region was found between variants recovered from chimpanzees, compared to levels of 10.4% nucleotide and 1.9% amino acid diversity between variants of HGV/GBV-C infecting humans (Adams *et al.* 1998). These findings suggested the existence of a species-specific association of GB viruses with their hosts, although the degree of sequence similarity contradicts with the high rate of sequence change of HGV/GBV-C estimated by Nakao (Nakao *et al.* 1997). In order to investigate this apparent paradox, a recent study examined the existence of restrictions on variability at certain sites in the HGV/GBV-C genome that may be responsible for the limited divergence of virus sequences over time (Simmonds and Smith,

1999). Potential RNA secondary structures detected and analysed by a covariance algorithm were found to be distributed throughout the HGV/GBV-C genome. Structural constraints imposed by the mechanism of RNA secondary structure formation may determine very slow, long term rate of sequence change in the HGV/GBV-C genome (Simmonds and Smith, 1999).

The identification of clusters of group-specific polymorphisms within the HGV/GBV-C 5'-UTR could improve the design of oligonucleotide primers for virus detection. In addition, the information could be used to improve the choice of restriction enzymes for the identification of HGV/GBV-C groups through RFLP typing of the 5'-UTR (Mukaide *et al.* 1997). This typing method, which has to be continuously modified to include new sequence data, can be used as a rapid and reliable technique to characterize new HGV/GBV-C variants isolated from different geographical areas.

Although phylogenetically, HGV/GBV-C has a genome organization similar to HCV and other members of *Flaviviridae* family (Leary *et al.* 1996; Linnen *et al.* 1996), it appears that this virus does not encode a core protein analogous to that of HCV. This possibility has been investigated further here by comparing the structure of HGV/GBV-C and HCV virions by ultracentrifugation (Chapter 4). HGV/GBV-C RNA was found to be present in single peaks at low density fractions within a range of 1.07 to 1.12 g/ml. This low density suggests an association between HGV/GBV-C particles and LDL, a prediction which was later confirmed in precipitation studies (Sato *et al.* 1996). The absence of viral fractions with densities higher than 1.17 g/ml is consistent with the results of

sequence analysis of the 5'-UTR suggesting that HGV/GBV-C lacks a nucleocapsid. The only AUG codon conserved amongst all twenty-four HGV/GBV-C sequences is the one near the amino-terminus of the E1 gene previously identified as the site of translation initiation by Simons *et al.* (Simons *et al.* 1996). An alternative possibility is that the HGV/GBV-C nucleocapsid is encoded by an open reading frame on the anti-genome strand (Kondo *et al.* 1998), although further studies are needed to confirm this.

The analysis of primary sequences proved to be a useful tool in determining structural features of the HGV/GBV-C 3'-UTR and predicting a consensus model for the secondary structure (Chapter 5). The high degree of nucleotide sequence conservation and the identification of a series of stable and conserved stem-loop structures within the HGV/GBV-C 3'-UTR suggest that this region is involved in RNA replication, similar to its function for other positive-strand RNA viruses (Brinton *et al.* 1986; Hahn *et al.* 1987; Deng and Brock, 1993; Kolykhalov *et al.* 1996; Blight and Rice, 1997). In particular, structural elements such as the long stable hairpin **VII** identified within the terminal part of the 3'-UTR have previously been shown to be particularly conserved amongst flaviviruses (Rice *et al.* 1985; Brinton *et al.* 1986; Proutski *et al.* 1997), pestiviruses (Deng and Brock, 1993), and HCV (Kolykhalov *et al.* 1996; Blight and Rice, 1997). Recently, experimental studies on flaviviruses have provided more information about the biological relevance of the predicted structures of the 3'-UTR of HCV (Yanagi *et al.* 1999) and BVDV (Yu *et al.* 1999). A series of HCV mutants containing large deletions in the 3'-UTR obtained from an

infectious clone by *in vitro* mutagenesis were tested sequentially for infectivity in a chimpanzee (Yanagi *et al.* 1999). The poly U/polypyrimidine region and the highly conserved region of 98 nucleotides at the 3'-terminal part which was predicted to form three stem-loop structures (Kolykhalov *et al.* 1996; Blight and Rice, 1997) proved to be critical for infectivity since mutants carrying deletions within these two regions of the HCV 3'-UTR failed to replicate in the chimpanzee.

In the case of BVDV, the predicted secondary structure of the 3'-UTR (Deng and Brock, 1993) was determined by RNase cleavage experiments combined with chemical modification of the native RNA transcripts (Yu *et al.* 1999). Largely in agreement with the proposed model, the conserved terminal part of the 3'-UTR was shown to contain two stem-loops (SLI and SLII) separated by a long single-stranded sequence (SS motif). Fifteen different mutations which affected either the 3' terminus of SLI or the SS motif were carried out in order to establish the functional role of these elements during RNA replication. Abolition of the stem regions of SLI and substitution of certain nucleotides within the SS motif totally destroyed the ability of the respective mutants to replicate in BHK cells.

In case of HGV/GBV-C, similar experimental approaches must be carried out to map the 3'-UTR structural elements predicted here (Chapter 5). Identifying and characterizing their functional significance by *in vitro* mutagenesis studies will require either an infectious clone as developed for HCV (Yanagi *et al.* 1998) or *in vitro* culture system as for BVDV (Yu *et al.* 1999). In

order to develop an efficient *in vitro* propagation system for HGV/GBV-C, the susceptibility of three types of cultured cells to virus infection was examined (Chapter 6). No evidence of viral replication in these systems was obtained under the experimental conditions tested. The highly stringent strand-specific RT-PCR assays performed here have previously been validated and used in a study which investigated whether extra-hepatic replication of HGV/GBV-C and HCV occurs in PBMCs (Mellor *et al.* 1998). While this *in vivo* study provided no evidence that PBMCs represent a replication site for HGV/GBV-C, confirming similar observations (Laskus *et al.* 1998; Radkowski *et al.* 1998), a more recent study suggested that HGV/GBV-C could infect and replicate in PBMCs *in vitro* (Fogeda *et al.* 1999). One HGV/GBV-C positive serum sample was used as inoculum in five different cell culture systems maintained for a 30 day-period (PBMCs isolated from four healthy donors and a cell pool obtained by mixing PBMCs from each donor). According to this study, intracellular plus-strand HGV/GBV-C RNA was detected soon after infection and remained present during the culture period while intracellular minus-strand HGV/GBV-C RNA appeared at day 7 p.i. and was detected either sporadically or continuously, depending on the different donors' PBMCs. However, these results remained to be confirmed in independent experiments while the identification of HGV/GBV-C replication sites remains an unresolved issue.

Considerable efforts have been made to improve current methodologies for the sensitive and specific detection of plus and minus strand RNA molecules. Problems related to false positive results have not been completely eliminated,

and results remain controversial. The protocols and conditions for the strand-specific detection of HCV and HGV/GBV-C RNAs described by various research groups need to be evaluated, compared and verified in order to establish reproducible assays with good sensitivity and specificity which can be widely used. Accurate and reliable detection assays for both plus- and minus-strand HGV/GBV-C RNAs will facilitate the identification of replication sites of this virus and the investigation of its possible pathogenicity.

Further investigations concerning the structure of HGV/GBV-C and whether or not the virus infection actually causes any disease are necessary to find the real answers. Although some reports support the idea that HGV/GBV-C is an 'innocent bystander' in comparison to the 'classical' agents of viral hepatitis, future research into the molecular virology of this virus should remain open to address basic questions such as: why HGV/GBV-C does not express epitopes, how it can survive without a core gene, for how long it has been in human populations.

REFERENCES

Aach, R. D., Szmuness, E., Mosley, J. W., Hollinger, B., Kahn, R. A., Stevens, C. E., Edwards, V. M., and Werch, J.(1981). Serum alanine aminotransferase of donors in relation to the risk on non-A, non-B hepatitis recipients: the transfusion-transmitted viruses study. *New England Journal of Medecine* **304**, 989-994.

Aach, R. D., Stevens, C. E., Hollinger, F. B., Mosley, J. W., Peterson, D. A., Taylor, P. E., Johnson, R. G., Barbosa, L. H., and Nemo, G. J.(1991). Hepatitis C virus infection in post-transfusion hepatitis. An analysis with first- and second-generation assays. *New England Journal of Medecine* **325**, 1325-1329.

Adams, N. J., Prescott, L. E., Jarvis, L. M., Lewis, J. C. M., Mcclure, M. O., Smith, D. B., and Simmonds, P.(1998). Detection of a novel flavivirus related to hepatitis G virus/GB virus C in chimpanzees. *Journal of General Virology* **79**, 1871-1877.

Almeida, J. D., Deinhardt, F., Holmes, A. W., Peterson, D. A., Wolfe, L., and Zuckerman, A. J.(1976). Morphology of the GB hepatitis agent. *Nature* **261**, 608-609.

Alter, H. J., Holland, P. V., Purcell, R. H., and Popper, H.(1978). Transmissible agent in non-A, non-B hepatitis. *The Lancet* **i**, 459-463.

Alter, H. J., Purcell, R. H., Shih, J. W., Melpolder, J. C., Houghton, M., Choo, Q. L., and Kuo, G.(1989). Detection of antibody to hepatitis C virus in prospectively followed transfusion recipients with acute and chronic non-A, non-B hepatitis. *New England Journal of Medecine* **321**, 1494-1500.

Alter, H. J., Nakatsuji, Y., Melpolder, J., Wages, J., Wesley, R., Shih, J. W. K., and Kim, J. P.(1997). The incidence of transfusion-associated hepatitis G virus infection and its relation to liver disease. *New England Journal of Medecine* **336**, 747-754.

An, P., Wei, L., Wu, X. Y., Yuhki, N., OBrien, S. J., and Winkler, C.(1997). Evolutionary analysis of the 5'-terminal region of hepatitis G virus isolated from different regions in China (Vol 78, pg 2477, 1997). *Journal of General Virology* **78**, 3371.

Aria, K. T. N., Sallie, R., Sangar, D., Alexander, G. J. M., Smith, H., Byrne, J., Portmann, B., Eddleston, A. L. W. F., and Williams, R.(1993). Detection of genomic and intermediate replicative strands of hepatitis C virus in liver tissue by insitu hybridization. *Journal of Clinical Investigation* **91**, 2226-2234.

Bartenschlager, R., Ahlbornlaake, L., Mous, J., and Jacobsen, H.(1993). Nonstructural protein-3 of the hepatitis C virus encodes a serine-type proteinase required for cleavage at the NS3/4 and NS4/5 junctions. *Journal of Virology* **67**,

3835-3844.

Belyaev, A. S., Chong, S., Novikov, A., Kongpachith, A., Masiarz, F. R., Lim, M., and Kim, J. P.(1998). Hepatitis G virus encodes protease activities which can effect processing of the virus putative nonstructural proteins. *Journal of Virology* **72**, 868-872.

Berg, T., Dirla, U., Naumann, U., Heuft, H. G., Kuther, S., Lobeck, H., Schreier, E., and Hopf, U.(1996). Responsiveness to interferon alpha treatment in patients with chronic hepatitis C coinfected with hepatitis G virus. *Journal of Hepatology* **25**, 763-768.

Blackwell, J. L. and Brinton, M. A.(1995). BHK Cell Proteins that Bind to the 3' Stem-Loop Structure of the West Nile Virus Genome RNA. *Journal of Virology* **69**, 5650-5658.

Blair, C. S., Davidson, F., Lycett, C., McDonald, D. M., Haydon, G. H., Yap, P. L., Hayes, P. C., Simmonds, P., and Gillon, J.(1998). Prevalence, incidence, and clinical characteristics of hepatitis G virus/GB virus C infection in Scottish blood donors. *Journal of Infectious Diseases* **178**, 1779-1782.

Blight, K., Trowbridge, R., Rowland, R., and Gowans, E.(1992). Detection of hepatitis-c virus RNA by insitu hybridization. *Liver* **12**, 286-289.

Blight, K., Rowland, R., Hall, P. D. M., Lesniewski, R. R., Trowbridge, R., LaBrooy, J. T., and Gowans, E. J.(1993). Immunohistochemical detection of the NS4 antigen of hepatitis C virus and its relation to histopathology. *American Journal of Pathology* **143**, 1568-1573.

Blight, K., Lesniewski, R. R., LaBrooy, J. T., and Gowans, E. J.(1994). Detection and distribution of hepatitis C-specific antigens in naturally infected liver. *Hepatology* **20**, 553-557.

Blight, K. J. and Rice, C. M.(1997). Secondary structure determination of the conserved 98-base sequence at the 3' terminus of hepatitis C virus genome RNA. *Journal of Virology* **71**, 7345-7352.

Bouffard, P., Hayashi, P. H., Acevedo, R., Levy, N., and Zeldis, J. B.(1992). Hepatitis C virus is detected in a monocyte macrophage subpopulation of peripheral blood mononuclear cells of infected patients. *Journal of Infectious Diseases* **166**, 1276-1280.

Boyer, J. C. and Haenni, A. L.(1994). Infectious Transcripts and cDNA clones of RNA Viruses. *Virology* **198**, 415-426.

Bradley, D., McCaustland, K., Krawczynski, K., Spelbring, J., Humphrey, C., and Cook, E. H.(1991). Hepatitis C virus: buoyant density of the factor VIII-derived

isolate in sucrose. *Journal of Medical Virology* **34**, 206-208.

Bradley, D. W., Maynard, J. E., Popper, H., Cook, E. H., Ebert, J. W., McCaustland, K. A., Schable, C. A., and Fields, H. A.(1983). Post-transfusion non-A, non-B hepatitis: physicochemical properties of two distinct agents. *Journal of Infectious Diseases* **148**, 254-265.

Bradley, D. W.(1990). Enterically-transmitted non-A, non-B hepatitis. *British Medical Bulletin* **46**, 442-461.

Bralet, M. P., Roudotthoraval, F., Pawlotsky, J. M., Bastie, A., Vannhieu, J. T., Duval, J., Dhumeaux, D., and Zafrani, E. S.(1997). Histopathologic impact of GB virus C infection on chronic hepatitis C. *Gastroenterology* **112**, 188-192.

Brandriss, M. W., Schlesinger, J. J., and Chapman, S. E.(1982). Growth of measles virus in a human macrophagelike cell line: U937. *American Journal of Pathology* **109**, 179-183.

Brinton, M. A., Fernandez, A. V., and Dispoto, J. H.(1986). The 3'-Nucleotides of Flavivirus Genomic RNA Form a Conserved Secondary Structure. *Virology* **153**, 113-121.

Brody, R. I., Eng, S., Melamed, J., Mizrachi, H., Schneider, R. J., Tobias, H., Teperman, L. W., and Theise, N. D.(1998). Immunohistochemical detection of hepatitis C antigen by monoclonal antibody TORDJI-22 compared with PCR viral detection. *American Journal of Clinical Pathology* **110**, 32-37.

Brown, E. A., Zhang, H. C., Ping, L. H., and Lemon, S. M.(1992). Secondary structure of the 5' nontranslated regions of hepatitis C virus and pestivirus genomic RNAs. *Nucleic Acids Research* **20**, 5041-5045.

Brown, K. E., Wong, S., Buu, M., Binh, T. V., Be, T. V., and Young, N. S.(1997a). High prevalence of GB virus C/hepatitis G virus in healthy persons in Ho Chi Minh City, Vietnam. *Journal of Infectious Diseases* **175**, 450-453.

Brown, K. E., Wong, S., and Young, N. S.(1997b). Prevalence of GBV-C/HGV, a novel 'hepatitis' virus, in patients with aplastic anaemia. *British Journal of Haematology* **97**, 492-496.

Bukh, J., Purcell, R. H., and Miller, R. H.(1994). Sequence analysis of the core gene of 14 hepatitis C virus genotypes. *Proceedings of the National Academy of Sciences, USA* **91**, 8239-8243.

Cha, T. A., Beall, E., Irvine, B., Kolberg, J., Chien, D., Kuo, G., and Urdea, M. S.(1992). At least 5 related, but distinct, hepatitis-C viral genotypes exist.

Proceedings of the National Academy of Sciences, USA **89**, 7144-7148.

Chan, S. W., McOmish, F., Holmes, E. C., Dow, B., Peutherer, J. F., Follett, E., Yap, P. L., and Simmonds, P.(1992). Analysis of a new hepatitis C virus type and its phylogenetic relationship to existing variants. *Journal of General Virology* **73**, 1131-1141.

Chaves, R. L., Graff, J., Normann, A., and Flehming, B.(1994). Specific detection of minus strand hepatitis A virus RNA by tail-PCR following reverse transcription. *Nucleic Acids Research* **22**, 1919-1920.

Chayama, K., Menon, J., Okamoto, K., Hashimoto, M., Tsubota, A., Kobayashi, M., Miyano, Y., Koike, H., Koida, I., Ikeda, K., and Kumada, H.(1997). Nucleotide sequences of hepatitis GB virus C. Identification of highly conserved domains in the 5' noncoding region and detection by polymerase chain reaction. *International Hepatology Communications* **6**, 225-232.

Chen, M., Sonnerborg, A., Johansson, B., and Sallberg, M.(1997). Detection of hepatitis G virus (GB virus C) RNA in human saliva. *Journal of Clinical Microbiology* **35**, 973-975.

Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., and Houghton, M.(1989). Isolation of a cDNA derived from a blood-borne non-A,

non-B hepatitis genome. *Science* **244**, 359-362.

Choo, Q. L., Richman, K. H., Han, J. H., Berger, K., Lee, C., Dong, C., Gallegos, C., Coit, D., Medina Selby, R., Barr, P. J., Weiner, A. J., Bradley, D. W., Kuo, G., and Houghton, M.(1991). Genetic organization and diversity of the hepatitis C virus. *Proceedings of the National Academy of Sciences, USA.* **88**, 2451-2455.

Conn, H. O. and Atterbury, C. E.(1993). *Diseases of the liver.* (Schiff, L. and Schiff, E. R. Eds.) J.B. Lippincott, Philadelphia. 875-934.

Cooper, R. A., Braunwald, A. D., and Kuo, A. L.(1982). Phorbol ester induction of leukemic cell differentiation is a membrane-mediated process. *Proceedings of the National Academy of Sciences, USA.* **79**, 2865-2869.

Cribier, B., Schmitt, C., Bingen, A., Kirn, A., and Keller, F.(1995). In vitro infection of peripheral blood mononuclear cells by hepatitis C virus. *Journal of General Virology* **76**, 2485-2491.

Dash, S., Halim, A. B., Tsuji, H., Hiramatsu, N., and Gerber, M. A.(1997). Transfection of HepG2 cells with infectious hepatitis C virus genome. *American Journal of Pathology* **151**, 363-373.

Davidson, F., Simmonds, P., Ferguson, J. C., Jarvis, L. M., Dow, B. C., Follett, E. A. C., Seed, C. R. G., Krusius, T., Lin, C., Medgyesi, G. A., Kiyokawa, H., Olim, G., Duraisamy, G., Cuypers, T., Saeed, A. A., Teo, D., Conradie, J., Kew, M. C., Lin, M., Nuchaprayoon, C., Ndimbie, O. K., and Yap, P. L.(1995). Survey of major genotypes and subtypes of hepatitis C virus using RFLP of sequences amplified from the 5' non- coding region. *Journal of General Virology* **76**, 1197-1204.

Dawson, G. J., Schlauder, G. G., PilotMatias, T. J., Thiele, D., Leary, T. P., Murphy, P., Rosenblatt, J. E., Simons, J. N., Martinson, F. E. A., Gutierrez, R. A., Lentino, J. R., Pachucki, C., Muerhoff, A. S., Widell, A., Tegtmeier, G., Desai, S., and Mushahwar, I. K.(1996). Prevalence studies of GB virus-C infection using reverse transcriptase polymerase chain reaction. *Journal of Medical Virology* **50**, 97-103.

Deinhardt, F., Holmes, A. W., Capps, R. P., and Popper, H.(1967). Studies on the transmission of human viral hepatitis to marmoset monkeys. *Journal of Experimental Medecine* **125**, 673-687.

Deka, N., Sharma, M. D., and Mukerjee, R.(1994). Isolation of the novel agent from human stool samples that is associated with sporadic non-A, non-B hepatitis. *Journal of Virology* **64**, 7810-7815.

Deng, R. T. and Brock, K. V.(1993). 5' and 3' untranslated regions of pestivirus genome - primary and secondary structure analyses. *Nucleic Acids Research* **21**, 1949-1957.

Diamantis, I., Bassetti, S., Erb, P., Ladewig, D., Gyr, K., and Battegay, M.(1997). High prevalence and coinfection rate of hepatitis G and C infections in intravenous drug addicts. *Journal of Hepatology* **26**, 794-797.

Dickson, R. C., Qian, K. P., and Lau, J. Y. N.(1997). High prevalence of GB virus-C/hepatitis G virus infection in liver transplant recipients. *Transplantation* **63**, 1695-1697.

Dille, B. J., Surowy, T. K., Gutierrez, R. A., Coleman, P. F., Knigge, M. F., Carrick, R. J., Aach, R. D., Hollinger, F. B., Stevens, C. E., Barbosa, L. H., Nemo, G. J., Mosley, J. W., Dawson, G. J., and Mushahwar, I. K.(1997). An ELISA for detection of antibodies to the E2 protein of GB virus C. *Journal of Infectious Diseases* **175**, 458-461.

Dusheiko, G., Schmilovitzweiss, H., Brown, D., McOmish, F., Yap, P. L., Sherlock, S., McIntyre, N., and Simmonds, P.(1994). Hepatitis C virus genotypes - an investigation of type- specific differences in geographic origin and disease. *Hepatology* **19**, 13-18.

Enomoto, M., Nishiguchi, S., Fukuda, K., Kuroki, T., Tanaka, M., Otani, S., Ogami, M., and Monna, T.(1998). Characteristics of patients with hepatitis C virus with and without GB virus C/hepatitis G virus co-infection and efficacy of interferon alfa. *Hepatology* **27**, 1388-1393.

Erker, J. C., Simons, J. N., Muerhoff, A. S., Leary, T. P., Chalmers, M. L., Desai, S. M., and Mushahwar, I. K.(1996). Molecular cloning and characterization of a GB virus C isolate from a patient with non-A-E hepatitis. *Journal of General Virology* **77**, 2713-2720.

Feinstone, S. M., Kapikian, A. Z., and Purcell, R. H.(1973). Hepatitis A: detection by immune electron microscopy of a viruslike antigen associated with acute illness. *Science* **182**, 1026-1028.

Feinstone, S. M., Kapikian, A. Z., and Purcell, R. H.(1975). Transfusion-associated hepatitis not due to viral hepatitis A or B. *New England Journal of Medicine* **292**, 767-770.

Feinstone, S. M., Mihalik, K., Kamimura, T., Alter, H. J., London, W. T., and Purcell, R. H.(1983). Inactivation of hepatitis B virus and non-A, non-B hepatitis by chloroform. *Infection and Immunity* **41**, 816-821.

Felsenstein, J.(1993). *PHYLIP Inference Package version 3.5*, Department of

Genetics, University of Washington, Seattle.

Feucht, H. H., Zollner, B., Polywka, S., and Laufs, R.(1996). Vertical transmission of hepatitis G. *Lancet* **347**, 615-616.

Feucht, H. H., Zollner, B., Polywka, S., Knodler, B., Schroter, M., Nolte, H., and Laufs, R.(1997). Prevalence of hepatitis G viraemia among healthy subjects, individuals with liver disease, and persons at risk for parenteral transmission. *Journal of Clinical Microbiology* **35** (**3**), 767-768.

Fiordalisi, G., Zanella, I., Mantero, G., Bettinardi, A., Stellini, R., Paraninfo, G., Cadeo, G., and Primi, D.(1996). High prevalence of GB virus C infection in a group of Italian patients with hepatitis of unknown aetiology. *Joural of Infectious Diseases* **174**, 181-183.

Fiordalisi, G., Bettinardi, A., Zanella, I., Stellini, R., Paraninfo, G., Cadeo, G., and Primi, D.(1997). Parenteral and sexual transmission of GB virus C and hepatitis C virus among human immunodeficient virus-positive patients. *Journal of Infectious Diseases* **175**, 1025-1026.

Fischler, B., Lara, C., Chen, M., Sonnerborg, A., Nemeth, A., and Sallberg, M.(1997). Genetic evidence for mother-to-infant transmission of hepatitis G virus. *Journal of Infectious Diseases* **176**, 281-285.

Fogeda, M., Navas, S., Martin, J., Casqueiro, M., Rodriguez, E., Arocena, C., and Carreno, V.(1999). *In vitro* infection of human peripheral blood mononuclear cells by GB virus C/hepatitis G virus. *Journal of Virology* **73**, 4052-4061.

Fournier, C., Sureau, C., Coste, J., Ducos, J., Pageaux, G., Larrey, D., Domergue, J., and Maurel, P.(1998). *In vitro* infection of adult normal human hepatocytes in primary culture by hepatitis C virus. *Journal of General Virology* **79**, 2367-2374.

Fried, M. W., Khudyakov, Y. E., Smallwood, G. A., Cong, M. E., Nichols, B., Diaz, E., Siefert, P., Gutekunst, K., Gordon, R. D., Boyer, T. D., and Fields, H. A.(1997). Hepatitis G virus co-infection in liver transplantation recipients with chronic hepatitis C and nonviral chronic liver disease. *Hepatology* **25**, 1271-1275.

Fukushi, S., Kurihara, C., Ishiyama, N., Okamura, H., Hoshino, F. B., Oya, A., and Katayama, K.(1996). Nucleotide sequence of the 5' noncoding region of hepatitis G virus isolated from Japanese patients: Comparison with reported isolates. *Biochemical and Biophysical Research Communications* **226**, 314-318.

Gabrielli, A., Manzin, A., Candela, M., Caniglia, M. L., Paolucci, S., Danieli, M. G., and Clementi, M.(1994). Active hepatitis C virus infection in bone marrow and peripheral blood mononuclear cells from patients with mixed

cryoglobulinaemia. *Clinical and Experimental Immunology* **97**, 87-93.

Galli, M., Zehender, G., Monti, G., Ballare, M., Saccardo, F., Piconi, S., Demaddalena, C., Bertoncelli, M. C., Rinaldi, G., Invernizzi, F., and Monteverde, A.(1995). Hepatitis C virus RNA in the bone marrow of patients with mixed cryoglobulinemia and in subjects with noncryoglobulinemic chronic hepatitis type C. *Journal of Infectious Diseases* **171**, 672-675.

Gastaldi, M., Massacrier, A., Planells, R., Robagliaschlupp, A., Portalbartolomei, I., Bourliere, M., Quilici, F., Fiteni, J., Mazzella, E., and Cau, P.(1995). Detection by in situ hybridization of hepatitis C virus positive and negative RNA strands using digoxigenin- labeled cRNA probes in human liver cells. *Journal of Hepatology* **23**, 509-518.

Gerber, M. A.(1995). Pathobiologic effects of hepatitis C. *Journal of Hepatology* **22**, 83-86.

Gonzalez-Peralta, R. P., Fang, J. W. S., Davis, G. L., Gish, R., Tsukiyama Kohara, K., Kohara, M., Mondelli, M. U., Lesniewski, R., Phillips, M. I., Mizokami, M., and Lau, J. Y. N.(1994). Optimization for the detection of hepatitis C virus antigens in the liver. *Journal of Hepatology* **20**, 143-147.

Grakoui, A., Mccourt, D. W., Wychowski, C., Feinstone, S. M., and Rice, C.

M.(1993). A second hepatitis C virus-encoded proteinase. *Proceedings of the National Academy of Sciences, USA* **90**, 10583-10587.

Gunji, T., Kato, N., Hijikata, M., Hayashi, K., Saitoh, S., and Shimotohno, K.(1994). Specific detection of positive and negative stranded hepatitis C viral RNA using chemical RNA modification. *Archives of Virology* **134**, 293-302.

Gussow, D. and Clackson, T.(1990). Direct clone characterisation from plaques and colonies by the polymerase chain reaction. *Nucleic Acids Research* **17**, 4000

Gutierrez, R. A., Dawson, G. J., Knigge, M. F., Melvin, S. L., Heynen, C. A., Kyrk, C. R., Young, C. E., Carrick, R. J., Schlauder, G. G., Surowy, T. K., Dille, B. J., Coleman, P. F., Thiele, D. L., Lentino, J. R., Pachucki, C., and Mushahwar, I. K.(1997). Seroprevalence of GB virus C and persistence of RNA and antibody. *Journal of Medical Virology* **53**, 167-173.

Haagsma, E. B., Cuypers, H. T. M., Gouw, A. S. H., Sjerps, M. C., Huizenga, J. R., Slooff, M. J. H., and Jansen, P. L. M.(1997). High prevalence of hepatitis G virus after liver transplantation without apparent influence on long-term graft function. *Journal of Hepatology* **26**, 921-925.

Hahn, J. A., Hahn, Y. S., Rice, C. M., Lee, E. H., Dalgarno, A., Strauss, E. G., and Strauss, J. H.(1987). Conserved Elements in the 3' Untranslated Region of

Flavivirus RNAs and Potential Cyclization Sequences. *Journal of Molecular Biology* **198**, 33-41.

Hammel, P., Marcellin, P., Martinotpeignoux, M., Pham, B. N., Degott, C., Level, R., Lefort, V., Benhallem, A., Erlinger, S., and Benhamou, J. P.(1994). Etiology of chronic hepatitis in france: predominant role of hepatitis C virus. *Journal of Hepatology* **21**, 618-623.

Hanley, J. P., Jarvis, L. M., Hayes, P. C., Lee, A. J., Simmonds, P., and Ludlam, C. A.(1998). Patterns of hepatitis G viraemia and liver disease in haemophiliacs previously exposed to non-virus inactivated coagulation factor concentrates. *Thrombosis and Haemostasis* **79**, 291-295.

Haruna, Y., Hayashi, N., Hiramatsu, N., Takehara, T., Hagiwara, H., Sasaki, Y., Kasahara, A., Fusamoto, H., and Kamada, T.(1993). Detection of hepatitis C virus RNA in liver tissues by an insitu hybridization technique. *Journal of Hepatology* **18**, 96-100.

Havekes, L., van Hinsbergh, V., Kempen, H. J., and Emeis, J.(1983). The metabolism *in vitro* of human low-density lipoprotein by the human hepatoma cell line Hep G2. *Biochemical Journal* **214**, 951-958.

He, L.-F., Alling, D., Popkin, T., Shapiro, M., Alter, H. J., and Purcell, R.

H.(1987). Determining the Size of Non-A, Non-B Hepatitis Virus by Filtration. *Journal of Infectious Diseases* **156**, 636-640.

Healey, C. J., Smith, D. B., Walker, J. L., Holmes, E. C., Fleming, K. A., Chapman, R. W. G., and Simmonds, P.(1995). Acute hepatitis C infection after sexual exposure. *Gut* **36**, 148-150.

Heringlake, S., Osterkamp, S., Trautwein, C., Tillmann, H. L., Boker, K., Muerhoff, S., Mushahwar, I. K., Hunsmann, G., and Manns, M. P.(1996). Association between fulminant hepatic failure and a strain of GBV virus C. *Lancet* **348**, 1626-1629.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., and Shimotohno, K.(1991a). Hypervariable regions in the putative glycoprotein of hepatitis C virus. *Biochemical and Biophysical Research Communications* **175**, 220-228.

Hijikata, M., Kato, N., Ootsuyama, Y., Nakagawa, M., and Shimotohno, K.(1991b). Gene mapping of the putative structural region of the hepatitis C virus genome by in vitro processing analysis. *Proceedings of the National Academy of Sciences, USA.* **88**, 5547-5551.

Hijikata, M., Shimizu, Y. K., Kato, H., Iwamoto, A., Shih, J. W., Alter, H. J., Purcell, R. H., and Yoshikura, H.(1993). Equilibrium centrifugation studies of hepatitis C virus -evidence for circulating immune complexes. *Journal of Virology* **67**, 1953-1958.

Hijikata, M. and Mishiro, S.(1996). Circulating immune complexes that contain HCV but not GBV-C in co-infected hosts. *International Hepatology Communications* **5**, 339-344.

Hiramatsu, N., Hayashi, N., Haruna, Y., Kasahara, A., Fusamoto, H., Mori, C., Fuke, I., Okayama, H., and Kamada, T.(1992). Immunohistochemical detection of hepatitis C virus-infected hepatocytes in chronic liver disease with monoclonal antibodies to core, envelope amd NS3 regions of the hepatitis C virus genome. *Hepatology* **16**, 306-311.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., and Schuster, P.(1994). Vienna RNA package. *Chemical Monthly* **125**, 167-188.

Hoofnagle, J. H., Gerety, R. J., Tabor, E., Feinstone, S. M., Barker, L. F., and Purcell, R. H.(1977). Transmission of non-A, non-B hepatitis. *Annals of Internal Medicine* **87**, 14-20.

Iacovacci, S., Manzin, A., Barca, S., Sargiacomo, M., Serafino, A., Valli, M. B., Macioce, G., Hassan, H. J., Ponzetto, A., Clement, M., Peschle, C., and Carloni, G.(1997). Molecular characterization and dynamics of hepatitis C virus replication in human fetal hepatocytes infected in vitro. *Hepatology* **26**, 1328-1337.

Ikeda, M., Sugiyama, K., Mizutani, T., Tanaka, T., Shimotohno, K., and Kato, N.(1997). Hepatitis G virus replication in human cultured cells displaying susceptibility to hepatitis C virus infection. *Biochemical and Biophysical Research Communications* **235**, 505-508.

Ito, T. and Lai, M. M. C.(1997). Determination of the secondary structure of and cellular protein binding to the 3'-untranslated region of the hepatitis C virus RNA genome. *Journal of Virology* **71**, 8698-8706.

Jackson, R. J., Hunt, S. L., Gibbs, C. L., and Kaminski, A.(1994). Internal initiation of translation of picornavirus RNAs (review). *Molecular Biology Reports* **19**, 147-159.

Jarvis, L. M., Davidson, F., Hanley, J. P., Yap, P. L., Ludlam, C. A., and Simmonds, P.(1996). Infection with hepatitis G virus among recipients of plasma products. *Lancet* **348**, 1352-1355.

Kaito, M., Watanabe, S., Tsukiyamakohara, K., Yamaguchi, K., Kobayashi, Y., Konishi, M., Yokoi, M., Ishida, S., Suzuki, S., and Kohara, M.(1994). Hepatitis C virus particle detected by immunoelectron microscopic study. *Journal of General Virology* **75**, 1755-1760.

Kamer, G. and Argos, P.(1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Research* **12**, 7269-7282.

Kanda, T., Yokosuka, O., Ehata, T., Maru, Y., Imazeki, F., Saisho, H., Shiratori, Y., and Omata, M.(1997). Detection of GBV-C RNA in patients with non-A-E fulminant hepatitis by reverse-transcription polymerase chain reaction. *Hepatology* **25**, 1261-1265.

Kanda, T., Yokosuka, O., Tagawa, M., Kawai, S., Imazeki, F., and Saisho, H.(1998). Quantitative analysis of GBV-C RNA in liver and serum by strand-specific reverse transcription-polymerase chain reaction. *Journal of Hepatology* **29**, 707-714.

Kanto, T., Hayashi, N., Takehara, T., Hagiwara, H., Mita, E., Naito, M., Kasahara, A., Fusamoto, H., and Kamada, T.(1995). Density analysis of hepatitis C virus particle population in the circulation of infected hosts: implications for virus neutralization or persistence. *Journal of Hepatology* **22**, 440-448.

Kao, J. H., Chen, P. J., and Chen, D. S.(1996a). GBV-C in the aetiology of fulminant hepatitis. *Lancet* **347**, 120

Kao, J. H., Chen, P. J., Hsiang, S. C., Chen, W., and Chen, D. S.(1996b). Phylogenetic analysis of GB virus C: comparison of isolates from Africa, North America, and Taiwan. *Journal of Infectious Diseases* **174**, 410-413.

Kao, J. H., Chen, P. J., Lai, M. Y., Chen, W., Liu, D. P., Wang, J. T., Shen, M. C., and Chen, D. S.(1997). GB virus-C/hepatitis G virus infection in an area endemic for viral hepatitis, chronic liver disease, and liver cancer. *Gastroenterology* **112**, 1265-1270.

Karayiannis, P., Petrovic, L. M., Fry, M., Moore, D., Enticott, M., McGarvey, M. J., Scheuer, P. J., and Thomas, H. C.(1989). Studies of GB hepatitis agent in tamarins. *Hepatology* **9**, 186-192.

Karayiannis, P., Hadziyannis, S. J., Kim, J., Pickering, J. M., Piatak, M., Hess, G., Yun, A., McGarvey, M. J., Wages, J., and Thomas, H. C.(1997a). hepatitis G virus infection: clinical characteristics and response to interferon. *Journal of Viral Hepatitis.* **4**, 37-44.

Karayiannis, P., Pickering, J., Chiaramonte, M., and Thomas, H. C.(1997b). Hepatitis G virus infection. *Lancet* **349**, 954.

Katayama, K., Fukushi, S., Kurihara, C., Ishiyama, N., Okamura, H., Hoshino, F. B., and Oya, A.(1997). New variant groups identified from HGV isolates. *Archives of Virology* **142**, 1021-1028.

Katayama, K., Kageyama, T., Fukushi, S., Hoshino, F. B., Kurihara, C., Ishiyama, N., Okamura, H., and Oya, A.(1998). Full-length GBV-C/HGV genomes from nine Japanese isolates: characterization by comparative analyses. *Archives of Virology* **143**, 1063-1075.

Katayama, Y., Apichartpiyakul, C., Handajani, R., Ishido, S., and Hotta, H.(1997). GB virus C hepatitis G virus (GBV-C/HGV) infection in Chiang Mai, Thailand, and identification of variants on the basis of 5'-untranslated region sequences. *Archives of Virology* **142**, 2433-2445.

Kato, N., Hijikata, M., Ootsuyama, Y., Nakagawa, M., Ohkoshi, S., Sugimura, T., and Shimotohno, K.(1990). Molecular cloning of the human hepatitis C virus genome from Japanese patients with non-A, non-B hepatitis. *Proceedings of the National Academy of Sciences, USA.* **87**, 9524-9528.

Kato, N., Nakazawa, T., Mizutani, T., and Shimotohno, K.(1995). Susceptibility of human t-lymphotropic virus type i infected cell line MT-2 to hepatitis C virus infection. *Biochemical and Biophysical Research Communications* **206**, 863-869.

Kato, T., Mizokami, M., Nakano, T., Orito, E., Ohba, K., Kondo, Y., Tanaka, Y., Ueda, R., Mukaide, M., Fujita, K., Yasuda, K., and Iino, S.(1998). Heterogeneity in E2 region of GBV-C/hepatitis G virus and hepatitis C virus. *Journal of Medical Virology* **55**, 109-117.

Kay, G. E., Lane, B. C., and Snyderman, D. R.(1983). Induction of selective biological responses to chemoattractants in a human monocyte-like cell line. *Infection and Immunity* **41**, 1166-1174.

Kodali, V. P., Gordon, S. C., Silverman, A. L., and McCray, D. G.(1994). Cryptogenic liver disease in the United States: further evidence for non-A, non-B, and non-C hepatitis. *American Journal of Gastroenterology* **89**, 1836-1839.

Kojima, M., Osuga, T., Tsuda, F., Tanaka, T., and Okamoto, H.(1994). Influence of antibodies to the hypervariable region of e2/NS1 glycoprotein on the selective replication of hepatitis C virus in chimpanzees. *Virology* **204**, 665-672.

Kolykhalov, A. A., Feinstone, S. M., and Rice, C. M.(1996). Identification of a highly conserved sequence element at the 3' terminus of hepatitis C virus genome RNA. *Journal of Virology* **70**, 3363-3371.

Kondo, Y., Mizokami, M., Nakano, T., Kato, T., Tanaka, Y., Hirashima, N., Ueda, R., Kunimatsu, M., Sasaki, M., Yasuda, K., and Iino, S.(1998). Analysis

of Conserved Ambisense Sequences within GB Virus C. *Journal of Infectious Diseases* **178**, 1185-1188.

Koonin, E. V.(1991). The phylogeny of RNA-dependent RNA polymerases of positive-strand RNA viruses. *Journal of General Virology* **72**, 2197-2206.

Koziol, D. E., Holland, P. V., Alling, D., Melpolder, J. C., Solomon, R. E., Purcell, R. H., Hudson, L. M., Shoup, F. J., Krakauer, H., and Alter, H. J.(1986). Antibody to hepatitis B core antigen as a paradoxical marker for non-A, non-B hepatitis agents in donated blood. *Annals of Internal Medicine* **104**, 488-495.

Kudo, T., Morishima, T., and Shibata, M.(1997). Hepatitis G infection. *New England Journal of Medicine* **337**, 276-277.

Kumar, S., Tamura, K., and Nei, M.(1993). *MEGA: Molecular Evolutionary Genetics Analysis, version 1.0*, Pennsylvania State University, Pennsylvania.

Kuo, G., Choo, Q. L., Alter, H. J., Gitnick, G. L., Redeker, A. G., Purcell, R. H., Miyamura, T., Dienstag, J. L., Alter, M. J., Stevens, C. E., Tegtmeier, F., Bonino, F., Columbo, M., Lee, W.-S., Kuo, C., Berger, K., Schuster, J. R., Overby, L. R., Bradley, D. W., and Houghton, M.(1989). An assay for circulating antibodies to a major etiologic virus of human non-A, non-B hepatitis. *Science*

244, 362-364.

Lamas, E., Baccarini, P., Housset, C., Kremsdorf, D., and Brechot, C.(1992). Detection of hepatitis C virus (HCV) RNA sequences in liver tissue by insitu hybridization. *Journal of Hepatology* **16**, 219-223.

Lanford, R. E., Sureau, C., Jacob, J. R., White, R., and Fuerst, T. R.(1994). Demonstration of in vitro infection of chimpanzee hepatocytes with hepatitis C virus using strand-specific RT/PCR. *Virology* **202**, 606-614.

Lanford, R. E., Chavez, D., Vonchisari, F., and Sureau, C.(1995). Lack of detection of negative-strand hepatitis C virus RNA in peripheral blood mononuclear cells and other extrahepatic tissues by the highly strand-specific rtth reverse transcriptase PCR. *Journal of Virology* **69**, 8079-8083.

Laskus, T., Radkowski, M., Wang, L. F., Cianciara, J., Vargas, H., and Rakela, J.(1997a). Hepatitis C virus negative strand RNA is not detected in peripheral blood mononuclear cells and viral sequences are identical to those in serum: a case against extrahepatic replication. *Journal of General Virology* **78**, 2747-2750.

Laskus, T., Radkowski, M., Wang, L. F., Vargas, H., and Rakela, J.(1997b). Lack of evidence for hepatitis G virus replication in the livers of patients coinfected with hepatitis C and G viruses. *Journal of Virology* **71**, 7804-7806.

Laskus, T., Radkowski, M., Wang, L.-F., Vargas, H., and Rakela, J.(1998a). Search for hepatitis C virus extrahepatic replication sites in patients with acquired immunodeficiency syndrome: specific detection of negative-strand viral RNA in various tissues. *Hepatology* **28**, 1398-1401.

Laskus, T., Radkowski, M., Wang, L. F., Vargas, H., and Rakela, J.(1998b). Detection of hepatitis G virus replication sites by using highly strand-specific Tth-based reverse transcriptase PCR. *Journal of Virology* **72**, 3072-3075.

Lau, G. K. K., Fang, J. W. S., Davis, G. L., and Lau, J. Y. N.(1994). Detection of hepatitis C virus genome in formalin-fixed paraffin-embedded liver tissue by *in situ* Reverse Transcription Polymerase Chain Reaction. *Journal of Medical Virology* **44**, 406-409.

Leary, T. P., Muerhoff, A. S., Simons, J. N., Pilot-Matias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M., and Mushahwar, I. K.(1996a). Consensus oligonucleotide primers for the detection of GB virus C in human cryptogenic hepatitis. *Journal of Virological Methods* **56**, 119-121.

Leary, T. P., Muerhoff, A. S., Simons, J. N., Pilot-Matias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. S., Dawson, G. J., Desai, S. M., and Mushahwar, I. K.(1996b). Sequence and genomic oganisation of GBV-C: a novel member of the Flaviviridae associated with human non-A-E hepatitis. *Journal of Medical*

Virology **48**, 60-67.

Leary, T. P., Muerhoff, A. S., Simons, J. N., PilotMatias, T. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Dawson, G. J., Desai, S. M., and Mushahwar, I. K.(1996c). Sequence and genomic organization of GBV-C: A novel member of the flaviviridae associated with human non-A-E hepatitis. *Journal of Medical Virology* **48**, 60-67.

Lerat, H., Berby, F., Trabaud, M. A., Vidalin, O., Major, M., Trepo, C., and Inchauspe, G.(1996). Specific detection of hepatitis C virus minus strand RNA in hematopoietic cells. *Journal of Clinical Investigation* **97**, 845-851.

Lesniewski, R., Okasinski, G., Carrick, R., Vansant, C., Desai, S., Johnson, R., Scheffel, J., and Moore, B.(1995). Antibody to hepatitis C virus second envelope (HCV-e2) glycoprotein: a new marker of HCV infection closely associated with viremia. *Journal of Medical Virology* **45**, 415-422.

Lewis, R. M., Cosgriff, T. M., Peters, C. J., and Morrill, J. C.(1987). Differentiation of a human monocytic cell line associated with increased production of Rift Valley Fever virus by infected cells. *Journal of Medical Virology* **23**, 207-215.

Lim, M. Y., Fry, K., Yun, A., Chong, S., Linnen, J., Fung, K., and Kim, J.

P.(1997). Sequence variation and phylogenetic analysis of envelope glycoprotein of hepatitis G virus. *Journal of General Virology* **78**, 2771-2777.

Lin, C., Pragai, B. M., Grakoui, A., Xu, J., and Rice, C. M.(1994). Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics. *Journal of Virology* **68**, 8147-8157.

Linnavuori, K. and Hovi, T.(1981). Herpes simplex virus infection in human monocyte cultures: Dose-dependent inhibition of monocyte differentiation resulting in abortive infection. *Journal of General Virology* **52**, 381-385.

Linnen, J., Wages, J., Zhang-Keck, Z., Fry, K. E., Krawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., and et al.(1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**, 505-508.

Lisitsyn, N. and Wigler, M.(1993). Cloning the differences between two complex genomes. *Science* **259**, 946-951.

LopezAlcorocho, J. M., Castillo, I., Tomas, J. F., and Carreno, V.(1999). Identification of a novel GB type C virus/hepatitis G virus subtype in patients with hematologic malignancies. *Journal of Medical Virology* **57**, 80-84.

Lou, S., Qiu, X. X., Tegtmeier, G., Leitza, S., Brackett, J., Cousineau, K., Varma, A., Seballos, H., Kundu, S., Kuemmerle, S., and Hunt, J. C.(1997). Immunoassays to study prevalence of antibody against GB virus C in blood donors. *Journal of Virological Methods* **68**, 45-55.

Luby, J. P., Burnett, W., and Hull, A. R.(1974). Relationship between cytomegalovirus and hepatic function abnormalities in the period after renal transplant. *Journal of Infectious Diseases* **129**, 511-518.

Madejon, A., Fogeda, M., Bartolome, J., Pardo, M., Gonzalez, C., Cotonat, T., and Carreno, V.(1997). GB virus C RNA in serum, liver, and peripheral blood mononuclear cells from patients with chronic hepatitis B, C, and D. *Gastroenterology* **113**, 573-578.

Manabe, S., Fuke, I., Tanishita, O., Kaji, C., Gomi, Y., Yoshida, S., Mori, C., Takamizawa, A., Yosida, I., and Okayama, H.(1994). Production of nonstructural proteins of hepatitis C virus requires a putative viral protease encoded by NS3. *Virology* **198**, 636-644.

Masuko, K., Mitsui, T., Iwano, K., Yamazaki, C., Okuda, K., Meguro, T., Murayama, N., Inoue, T., Tsuda, F., Okamoto, H., Miyakawa, Y., and Mayumi, M.(1996). Infection with hepatitis GB virus C in patients on maintenance hemodialysis. *New England Journal of Medicine* **334**, 1485-1490.

Matsumoto, M., Hwang, S. B., Jeng, K. S., Zhu, N., and Lai, M. M. C.(1996). Homotypic interaction and multimerization of hepatitis C virus core protein. *Virology* **218**, 43-51.

Matzura, O. and Wennborg, A.(1996). RNAdraw: An integrated program for RNA secondary structure calculation and analysis under 32-bit Microsoft Windows. *Computer Applications in Bioscience* **12**, 247-249.

Mcguinness, P. H., Bishop, G. A., Mccaughan, G. W., Trowbridge, R., and Gowans, E. J.(1994). False detection of negative-strand hepatitis C virus RNA. *Lancet* **343**, 551-552.

McOmish, F., Chan, S.-W., Dow, B. C., Gillon, J., Frame, W. D., Crawford, R. J., Yap, P. L., Follett, E. A. C., and Simmonds, P.(1993). Detection of three types of hepatitis C virus in blood donors: Investigation of type-specific differences in serological reactivity and rate of alanine aminotransferase abnormalities. *Transfusion* **33**, 7-13.

McOmish, F., Yap, P. L., Dow, B. C., Follett, E. A. C., Seed, C., Keller, A. J., Cobain, T. J., Krusius, T., Kolho, E., Naukkarinen, R., Lin, C., Lai, C., Leong, S., Medgyesi, G. A., Hejjas, M., Kiyokawa, H., Fukada, K., Cuypers, T., Saeed, A. A., Alrasheed, A. M., Lin, M., and Simmonds, P.(1994). Geographical distribution of hepatitis C virus genotypes in blood donors - an international

collaborative survey. *Journal of Clinical Microbiology* **32**, 884-892.

Mellor, J., Holmes, E. C., Jarvis, L. M., Yap, P. L., Simmonds, P., and International Collaborators(1995). Investigation of the pattern of hepatitis C virus sequence diversity in different geographical regions: implications for virus classification. *Journal of General Virology* **76**, 2493-2507.

Mellor, J., Walsh, E. A., Prescott, L. E., Jarvis, L. M., Davidson, F., Yap, P. L., Simmonds, P., Nowicki, M. J., Mosley, J. W., Lin, C. K., Lai, C. L., deOlim, G., Martins, I. A., Ong, Y. W., Teo, D., Lin, M., Nuchprayoon, C., and Tanprasert, S.(1996). Survey of type 6 group variants of hepatitis C virus in southeast Asia by using a core-based genotyping assay. *Journal of Clinical Microbiology* **34**, 417-423.

Mellor, J., Haydon, G. H., Blair, C., Livingstone, W. J., and Simmonds, P.(1998). Low level or absent in vivo replication of hepatitis C virus and hepatitis G virus/GB virus C in peripheral blood mononuclear cells. *Journal of General Virology* **79**, 705-714.

Melnick, J. L.(1982). Classification of hepatitis A virus as enterovirus type 72 and of hepatitis B virus as hepadnovirus type 1. *Intervirology* **18**, 105-106.

Melvin, S. L., Dawson, G. J., Carrick, R. J., Schlauder, G. G., Heynen, C. A.,

and Mushahwar, I. K.(1998). Biophysical characterization of GB virus C from human plasma. *Journal of Virological Methods* **71**, 147-157.

Miller, R. H. and Purcell, R. H.(1990). Hepatitis C virus shares amino acid sequence similarity with pestiviruses and flaviviruses as well as members of two plant virus supergroups. *Proceedings of the National Academy of Sciences, USA.* **87**, 2057-2061.

Miyamoto, H., Okamoto, H., Sato, K., Tanaka, T., and Mishiro, S.(1992). Extraordinarily low density of hepatitis C virus estimated by sucrose density gradient centrifugation and the polymerase chain reaction. *Journal of General Virology* **73**, 715-718.

Miyamura, T. and Matsuura, Y.(1993). Structural proteins of hepatitis C virus. *Trends in Microbiology* **1**, 229-231.

Mizutani, T., Kato, N., Ikeda, M., Sugiyama, K., and Shimotohno, K.(1996). Long-term human T-cell culture system supporting hepatitis C virus replication. *Biochemical and Biophysical Research Communications* **227**, 822-826.

Moaven, L. D., Hyland, C. A., Young, I. F., Bowden, D. S., McCaw, R., Mison, L., and Locarnini, S. A.(1996a). Prevalence of hepatitis G virus in Queensland blood donors. *Medical Journal of Australia* **165**, 369-371.

Moaven, L. D., Tennakoon, P. S., Bowden, D. S., and Locarnini, S. A.(1996b). Mother-to-baby transmission of hepatitis G virus. *Medical Journal of Australia* **165**, 84-85.

Moldvay, J., Deny, P., Pol, S., Brechot, C., and Lamas, E.(1994). Detection of hepatitis C virus RNA in peripheral blood mononuclear cells of infected patients by in situ hybridization. *Blood* **83**, 269-273.

Muerhoff, A. S., Leary, T. P., Simons, J. N., Pilot-Matias, T. J., Dawson, G. J., Erker, J. C., Chalmers, M. L., Schlauder, G. G., Desai, S. M., and Mushahwar, I. K.(1995). Genomic Organization of GB Viruses A and B: Two New Members of the Flaviviridae Associated with GB Agent Hepatitis. *Journal of Virology* **69**, 5621-5630.

Muerhoff, A. S., Simons, J. N., Erker, J. C., Desai, S. M., and Mushahwar, I. K.(1996a). Identification of conserved nucleotide sequences within the GB virus C 5'-untranslated region: design of PCR primers for detection of viral RNA. *Journal of Virological Methods* **62**, 55-62.

Muerhoff, A. S., Simons, J. N., Leary, T. P., Erker, J. C., Chalmers, M. L., Pilot-Matias, T. J., Dawson, G. J., Desai, S. M., and Mushahwar, I. K.(1996b). Sequence heterogeneity within the 5'-terminal region of the hepatitis GB virus C genome and evidence for genotypes. *Journal of Hepatology* **25**, 379-384.

Muerhoff, A. S., Smith, D. B., Leary, T. P., Erker, J. C., Desai, S. M., and Mushahwar, I. K.(1997). Identification of GB virus C variants by phylogenetic analysis of 5'-untranslated and coding region sequences. *Journal of Virology* **71**, 6501-6508.

Mukaide, M., Mizokami, M., Orito, E., Ohba, K., Nakano, T., Ueda, R., Hikiji, K., Iino, S., Shapiro, S., Lahat, N., Park, Y. M., Kim, B. S., Oyunsuren, T., Rezieg, M., AlAhdal, M. N., and Lau, J. Y. N.(1997). Three different GB virus C/hepatitis G virus genotypes -Phylogenetic analysis and a genotyping assay based on restriction fragment length polymorphism. *FEBS Letters* **407**, 51-58.

Muller, H. M., Pfaff, E., Goeser, T., Kallinowski, B., Solbach, C., and Theilmann, L.(1993). Peripheral blood leukocytes serve as a possible extrahepatic site for hepatitis C virus replication. *Journal of General Virology* **74**, 669-676.

Nakajima, N., Hijikata, M., Yoshikura, H., and Shimizu, Y. K.(1996). Characterization of long-term cultures of hepatitis C virus. *Journal of Virology* **70**, 3325-3329.

Nakao, H., Okamoto, H., Fukuda, M., Tsuda, F., Mitsui, T., Masuko, K., Lizuka, H., Miyakawa, Y., and Mayumi, M.(1997). Mutation rate of GB virus C hepatitis G virus over the entire genome and in subgenomic regions. *Virology* **233**, 43-50.

Nei, M.(1987). *Molecular Evolutionary Genetics*, Columbia University Press,

Nubling, C. M. and Lower, J.(1996). GB-C genomes in a high-risk group, in plasma pools and in intravenous immunoglobulin. *The Lancet* **347**, 68

Nuovo, G. J., Lidonnici, K., Macconnell, P., and Lane, B.(1993). Intracellular localization of polymerase chain reaction (PCR)-amplified hepatitis C cDNA. *American Journal of Surgical Pathology* **17**, 683-690.

O'Neill, R. E. and Palese, P.(1994). Cis-acting signals and trans-acting factors involved in influenza virus RNA synthesis. *Infectious Agents and Diseases* **3**, 77-84.

Okamoto, H., Nakao, H., Inoue, T., Fukuda, M., Kishimoto, J., Iizuka, H., Tsuda, F., Miyakawa, Y., and Mayumi, M.(1997). The entire nucleotide sequences of two GB virus C/hepatitis G virus isolates of distinct genotypes from Japan. *Journal of General Virology* **78**, 737-745.

Orito, E., Mizokami, M., Nakano, T., Wu, R. R., Cao, K., Ohba, K. I., Ueda, R., Mukaide, M., Hikiji, K., Matsumoto, Y., and Iino, S.(1996). GB virus C hepatitis G virus infection among Japanese patients with chronic liver diseases and blood donors. *Virus Research* **46**, 89-93.

Pawlotsky, J. M., Roudotthoraval, F., Muerhoff, A. S., Pellerin, M., Germanidis, G., Desai, S. M., Bastie, A., Darthuy, F., Remire, J., Zafrani, E. S., Soussy, C. J., Mushahwar, I. K., and Dhumeaux, D.(1998). GB virus C (GBV-C) infection in patients with chronic hepatitis C. Influence on liver disease and on hepatitis virus behaviour: Effect of interferon alfa therapy. *Journal of Medical Virology* **54**, 26-37.

Pessoa, M. G., Terrault, N. A., Ferrell, L. D., Kim, J. P., Kolberg, J., Detmer, J., Collins, M. L., Yun, A. J., Viele, M., Lake, J. R., Roberts, J. P., Ascher, N. L., and Wright, T. L.(1997). Hepatitis G virus in patients with cryptogenic liver disease undergoing liver transplantation. *Hepatology* **25**, 1266-1270.

Pickering, J. M., Thomas, H. C., and Karayiannis, P.(1997a). Genetic diversity between hepatitis G virus isolates: Analysis of nucleotide variation in the NS-3 and putative 'core' peptide genes. *Journal of General Virology* **78**, 53-60.

Pickering, J. M., Thomas, H. C., and Karayiannis, P.(1997b). Predicted secondary structure of the hepatitis G virus and GB virus-A 5' untranslated regions consistent with an internal ribosome entry site. *Journal of Viral Hepatitis.* **4**, 175-184.

PilotMatias, T. J., Muerhoff, A. S., Simons, J. N., Leary, T. P., Buijk, S. L., Chalmers, M. L., Erker, J. C., Dawson, G. J., Desai, S. M., and Mushahwar, I.

K.(1996). Identification of antigenic regions in the GB hepatitis viruses GBV-A GBV-B, and GBV-C. *Journal of Medical Virology* **48**, 329-338.

Polish, L. B., Gallagher, M., Fields, H. A., and Hadler, S. C.(1993). Delta hepatitis: molecular biology and clinical and epidemiological features. *Clinical Microbiology Reviews* **6**, 211-229.

Poole, T. L., Wang, C. Y., Popp, R. A., Potgieter, L. N. D., Siddiqui, A., and Marc, S.(1995). Pestivirus translation initiation occurs by internal ribosome entry. *Virology* **206**, 750-754.

Prince, A. M., Brotman, B., Grady, G. F., Kuhns, W. J., Hazzi, C., Levine, R. W., and Millian, S. J.(1974). Long-incubation post-transfusion hepatitis without serological evidence of exposure to hepatitis B virus. *The Lancet* **ii**, 241-246.

Proutski, V., Gould, E. A., and Holmes, E. C.(1997). Secondary structure of the 3' untranslated region of flaviviruses: similarities and differences. *Nucleic Acids Research* **25**, 1194-1202.

Radkowski, M., Wang, L. F., Vargas, H., Rakela, J., and Laskus, T.(1998). Lack of evidence for GB virus C/hepatitis G virus replication in peripheral blood mononuclear cells. *Journal of Hepatology* **28**, 179-183.

Reyes, G. R., Purdy, M., Kim, J. S., Luk, K. C., Young, L. M., Fry, K. E., and Bradley, D. W.(1990). Isolation of cDNA from the virus responsible for enterically transmitted non-A, non-B hepatitis. *Science* **247**, 1335-1339.

Rice, C. M., Lenches, E. M., Eddy, S. R., Shin, S. J., Sheets, R. L., and Strauss, J. H.(1985). Nucleotide sequence of yellow fever virus: implications for flavivirus gene expression and evolution. *Science* **229**, 726-733.

Rigby, W. F. C., Shen, L., Ball, E. D., Guyre, P. M., and Fanger, M. W.(1984). Differentiation of a human monocytic cell line by 1, 25-dihydroxy vitamin D_3 (calcitriol): a morphologic, phenotypic, and functional analysis. *Blood* **64**, 1110-1115.

Rizzetto, M., Canese, M. G., Arico, S., Crivelli, O., Trepo, C., Bonino, F., and Verme, G.(1977). Immunofluorescence detection of new antigen-antibody system (δ/anti-δ) associated to hepatitis B virus in liver and in serum of HBsAg carriers. . *Gut* **18**, 997-1003.

Ross, R. S., Viazov, S., Schmitt, U., Schmolke, S., Tacke, M., OfenlochHaehnle, B., Holtmann, M., Muller, N., DaVilla, G., Yoshida, C. F., Oliveira, J. M., Szabo, A., Paladi, N., Kruppenbacher, J. P., Philipp, T., and Roggendorf, M.(1998). Distinct prevalence of antibodies to the E2 protein of GB virus C/ hepatitis G virus in different parts of the world. *Journal of Medical Virology* **54**,

103-106.

Rubio, A., Rey, C., Sanchezquijano, A., Leal, M., Pineda, J. A., Lissen, E., and Hess, G.(1997). Is hepatitis G virus transmitted sexually? *JAMA* **277**, 532-533.

Saito, S., Tanaka, K., Kondo, M., Morita, K., Kitamura, T., Kiba, T., Numata, K., and Sekihara, H.(1997). Plus- and minus-stranded hepatitis G virus RNA in liver tissue and in peripheral blood mononuclear cells. *Biochemical and Biophysical Research Communications* **237**, 288-291.

Saitou, N. and Nei, M.(1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.

Saiz, J. C., Ampurdanes, S., Olmedo, E., LopezLabrador, F. X., Forns, X., Guilera, M., Tassies, D., Costa, J., Sancheztapias, J. M., deAnta, M. T. J., and Rodes, J.(1997). Hepatitis G virus infection in chronic hepatitis C: Frequency, features and response to interferon therapy. *Journal of Hepatology* **26**, 787-793.

Sangar, D. V. and Carroll, A. R.(1998). A tale of two strands: Reverse-Transcriptase Polymerase Chain Reaction detection of hepatitis C virus replication. *Hepatology* **28**, 1173-1176.

Santolini, E., Migliaccio, G., and Lamonica, N.(1994). Biosynthesis and

biochemical properties of the hepatitis C virus core protein. *Journal of Virology* **68**, 3631-3641.

Sarrazin, C., Herrmann, G., Roth, W. K., Lee, J. H., Marx, S., and Zeuzem, S.(1997). Prevalence and clinical and histological manifestation of hepatitis G/GBV-C infections in patients with elevated aminotransferases of unknown etiology. *Journal of Hepatology* **27**, 276-283.

Sato, K., Tanaka, T., Okamoto, H., Miyakawa, Y., and Mayumi, M.(1996). Association of circulating hepatitis G virus with lipoproteins for a lack of binding with antibodies. *Biochemical and Biophysical Research Communications* **229**, 719-725.

Scallan, M. F., Clutterbuck, D., Jarvis, L. M., Scott, G. R., and Simmonds, P.(1998). Sexual transmission of GB virus-C/hepatitis G virus. *Journal of Medical Virology* **55**, 203-208.

Schlauder, G. G., Dawson, G. J., Simons, J. N., Pilot-Matias, T. J., Guiterrez, R. A., Heynen, C. A., Knigge, M. F., Kurpieqski, G. S., Buijk, S. L., Leary, T. P., Muerhoff, A. S., Desai, S. M., and Mushahwar, I. K.(1995). Molecular and serologic analysis in the transmission of the GB hepatitis agents. *Journal of Medical Virology* **46**, 81-90.

Schlueter, V., Schmolke, S., Stark, K., Hess, G., OfenlochHaehnle, B., and Engel, A. M.(1996). Reverse transcription-PCR detection of hepatitis G virus. *Journal of Clinical Microbiology* **34**, 2660-2664.

Schreier, E., Hohne, M., Kunkel, U., Berg, T., and Hopf, U.(1996). Hepatitis GBV-C sequences in patients infected with HCV contaminated anti-D immunoglobulin and among iv drug users in Germany. *Journal of Hepatolology* **25**, 385-389.

Seipp, S., Mueller, H. M., Pfaff, E., Stremmel, W., Theilmann, L., and Goeser, T.(1997). Establishment of persistent hepatitis C virus infection and replication in vitro. *Journal of General Virology* **78**, 2467-2476.

Sheng, L., Soumillion, A., Peerlinck, K., Verslype, C., Lin, L., vanPelt, J., Hess, G., Vermylen, J., and Yap, S. H.(1997). Hepatitis G viral RNA in serum and in peripheral blood mononuclear cells and its relation to HCV-RNA in patients with clotting disorders. *Thrombosis and Haemostasis* **77**, 868-872.

Shimizu, Y. K., Iwamoto, A., Hijikata, M., Purcell, R. H., and Yoshikura, H.(1992). Evidence for in vitro replication of hepatitis C virus genome in a human T-cell line. *Proceedings of the National Academy of Sciences, USA.* **89**, 5477-5481.

Simmonds, P., Balfe, P., Peutherer, J. F., Ludlam, C. A., Bishop, J. O., and Brown, A. J.(1990a). Human immunodeficiency virus-infected individuals contain provirus in small numbers of peripheral mononuclear cells and at low copy numbers. *Journal of Virology* **64**, 864-872.

Simmonds, P., Zhang, L. Q., Watson, H. G., Rebus, S., Ferguson, E. D., Balfe, P., Leadbetter, G. H., Yap, P. L., Peutherer, J. F., and Ludlam, C. A.(1990b). Hepatitis C quantification and sequencing in blood products, haemophiliacs, and drug users. *Lancet* **336**, 1469-1472.

Simmonds, P., Holmes, E. C., Cha, T. A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J., and Urdea, M. S.(1993a). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *Journal of General Virology* **74**, 2391-2399.

Simmonds, P., McOmish, F., Yap, P. L., Chan, S. W., Lin, C. K., Dusheiko, G., Saeed, A. A., and Holmes, E. C.(1993b). Sequence variability in the 5' non coding region of hepatitis C virus: identification of a new virus type and restrictions on sequence diversity. *Journal of General Virology* **74**, 661-668.

Simmonds, P., Smith, D. B., McOmish, F., Yap, P. L., Kolberg, J., Urdea, M. S., and Holmes, E. C.(1994). Identification of genotypes of hepatitis C virus by

sequence comparisons in the core, E1 and NS-5 regions. *Journal of General Virology* **75**, 1053-1061.

Simmonds, P.(1995). Variability of hepatitis C virus. *Hepatology* **21**, 570-583.

Simmonds, P.(1997). Clinical relevance of hepatitis C virus genotypes. *Gut* **40**, 291-293.

Simmonds, P. and Smith, D. B.(1999). Structural constraints on RNA virus evolution. *Journal of Virology* **73**, 5787-5794.

Simons, J. N., Leary, T. P., Dawson, G. J., Pilot-Matias, T. J., Muerhoff, A. S., Schlauder, G. G., Desai, S. M., and Mushahwar, I. K.(1995a). Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine* **1**, 564-569.

Simons, J. N., Pilot-Matias, T. J., Leary, T. P., Dawson, G. J., Desai, S. M., Schlauder, G. G., Muerhoff, A. S., Erker, J. C., Buijk, S. L., Chalmers, M. L., Vansant, C. L., and Mushahwar, I. K.(1995b). Identification of two flavivirus-like genomes in the GB hepatitis agent. *Proceedings of the National Academy of Sciences, USA* **92**, 3401-3405.

Simons, J. N., Desai, S. M., Schultz, D. E., Lemon, S. M., and Mushahwar, I. K.(1996). Translation initation in GB viruses A and C: evidence for internal

ribosome entry and implications for genome organisation. *Journal of Virology* **70**, 6126-6135.

Smith, D. B., Mellor, J., Jarvis, L. M., Davidson, F., Kolberg, J., Urdea, M., Yap, P. L., Simmonds, P., Conradie, J. D., Neill, A. G. S., Dusheiko, G. M., Kew, M. C., Crookes, R., Koshy, A., Lin, C. K., Lai, C., Murraylyon, I. M., Elguneid, A., Gunaid, A. A., Yemen, T., Yemen, S., Mutimer, D., Ahmed, M., Nuchprayoon, C., Tanprasert, S., Preston, F. E., Makris, M., Chuansumrit, A., Mahasandana, C., Pritchard, D., Riley, E., Greenwood, B. M., Saeed, A. A., Alrasheed, A. M., Saleh, M. G., Mcfarlane, I., Tibbs, C., Williams, R., Power, J., Lawlor, E., and Kiyokawa, H.(1995). Variation of the hepatitis C virus 5' non-coding region: implications for secondary structure, virus detection and typing. *Journal of General Virology* **76**, 1749-1761.

Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P. L., and Simmonds, P.(1997). The origin of hepatitis C virus genotypes. *Journal of General Virology* **78**, 321-328.

Strauss, E. G. and Strauss, J. H.(1983). Replication strategies of the single stranded RNA viruses of Eukaryotes. *Current Topics in Microbiology and Immunology* **105**, 1-99.

Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van

Heuverswyn, H., and Maertens, G.(1993). Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *Journal of General Virology* **74**, 1093-1102.

Sugg, U., Schenzie, D., and Hess, G.(1988). Antibodies to hepatitis B core antigen in blood donors screened for alanine aminotransferase level and hepatitis non-A, non-B in recipients. *Transfusion* **28**, 386-388.

Sundstrom, C. and Nilsson, K.(1976). Establishment and characterization of human histiocytic lymphoma cell line (U-937). *International Journal of Cancer* **17**, 565-577.

Surowy, T. K., Leary, T. P., Carrick, R. J., Knigge, M. F., Pilot-Matias, T. J., Heynen, C., Gutierrez, R. A., Desai, S. M., Dawson, G. J., and Mushahwar, I. K.(1997). GB virus C E2 glycoprotein: expression in CHO cells, purification and characterisation. *Journal of General Virology* **78**, 1851-1859.

Tacke, M., Kiyosawa, K., Stark, K., Schleuter, V., OfenlochHaehnle, B., Hess, G., and Engel, A. M.(1997a). Detection of antibodies to a putative hepatitis g virus envelope protein. *Lancet* **349**, 318-320.

Tacke, M., Schmolke, S., Schlueter, V., Sauleda, S., Esteban, J. I., Tanaka, E., Kiyosawa, K., Alter, H. J., Schmitt, U., Hess, G., OfenlochHaehnle, B., and

Engel, A. M.(1997b). Humoral immune response to the E2 protein of hepatitis G virus is associated with long-term recovery from infection and reveals a high frequency of hepatitis G virus exposure among healthy blood donors. *Hepatology* **26**, 1626-1633.

Takahashi, K., Hijikata, M., Aoyama, K., Hoshino, H., Hino, K., and Mishiro, S.(1997). Characterization of GBV-C/HGV viral genome: Comparison among different isolates for a similar to 2 kb-sequence that covers entire E1 and most of 5'UTR and E2. *International Hepatology Communications* **6**, 253-263.

Takamizawa, A., Mori, C., Fuke, I., Manabe, S., Murakami, S., Fujita, J., Onishi, E., Andoh, T., Yoshida, I., and Okayama, H.(1991). Structure and organization of the hepatitis C virus genome isolated from human carriers. *Journal of Virology* **65**, 1105-1113.

Tanaka, E., Alter, H. J., Nakatsuji, Y., Shih, J. W. K., Kim, J. P., Matsumoto, A., Kobayashi, M., and Kiyosawa, K.(1996). Effect of hepatitis G virus infection on chronic hepatitis C. *Annals of Internal Medicine* **125**, 740-743.

Tanaka, Y., Enomoto, N., Kojima, S., Tang, L., Goto, M., Marumo, F., and Sato, C.(1993). Detection of hepatitis C virus RNA in the liver by *in situ* hybridization. *Liver* **13**, 203-208.

Tanaka, Y., Mizokami, M., Orito, E., Ohba, K., Kato, T., Kondo, Y., Mboudjeka, I., Zekeng, L., Kaptue, L., Bikandou, B., MPele, P., Takehisa, J., Hayami, M., Suzuki, Y., and Gojobori, T.(1998). African origin of GB virus C hepatitis G virus. *FEBS Letters* **423**, 143-148.

Taniguchi, S., Okamoto, H., Sakamoto, M., Kojima, M., Tsuda, F., Tanaka, T., Munekata, E., Muchmore, E. E., Peterson, D. A., and Mishiro, S.(1993). A structurally flexible and antigenically variable n- terminal domain of the hepatitis C virus e2/NS1 protein - implication for an escape from antibody. *Virology* **195**, 297-301.

Tassopoulos, N. C., Hatzakis, A., Delladetsima, I., Koutelou, M. G., Todoulos, A., and Miriagou, V.(1992). Role of hepatitis C virus in acute non-A, non-B hepatitis in Greece: A 5-year prospective study. *Gastroenterology* **102**, 969-972.

Tassopoulos, N. C., Krawczynski, K., Hatzakis, A., Katsoulidou, A., Delladetsima, I., Koutelou, M. G., and Trichopoulos, D.(1994). Role of hepatitis e virus in the etiology of community- acquired non-A, non-B hepatitis in Greece - case report. *Journal of Medical Virology* **42**, 124-128.

Theodore, D. and Lemon, S. M.(1997). GB virus C, hepatitis G virus, or human orphan flavivirus? *Hepatology* **25**, 1285-1286.

Thomssen, R., Bonk, S., Propfe, C., Heermann, K. H., Kochel, H. G., and Uy, A.(1992). Association of hepatitis c virus in human sera with beta-lipoprotein. *Medical Microbiology and Immunology* **181**, 293-300.

Thomssen, R., Bonk, S., and Thiele, A.(1993). Density heterogeneities of hepatitis C virus in human sera due to the binding of beta-lipoproteins and immunoglobulins. *Medical Microbiology and Immunology* **182**, 329-334.

Tokita, H., Shrestha, S. M., Okamoto, H., Sakamoto, M., Hirokita, M., Iizuka, H., Shrestha, S., Miyakawa, Y., and Mayumi, M.(1994). Hepatitis C virus variants from Nepal with novel genotypes and their classification into the third major group. *Journal of General Virology* **75**, 931-936.

Tsuda, F., Hadiwandowo, S., Sawada, N., Fukuda, M., Tanaka, T., Okamoto, H., Miyakawa, Y., and Mayumi, M.(1996). Infection with GB virus C (GBV-C) in patients with chronic liver disease or on maintenance hemodialysis in Indonesia. *Journal of Medical Virology* **49**, 248-252.

Tsukiyama Kohara, K., Iizuka, N., Kohara, M., and Nomoto, A.(1992). Internal ribosome entry site within hepatitis C virus RNA. *Journal of Virology* **66**, 1476-1483.

Vargas, H. E., Laskus, T., Radkowski, M., Poutous, A., Wang, L. F., Lee, R.,

Dodson, F., Gayowski, T., Singh, N., Marino, I. R., Fung, J. J., ZhangKeck, Z. Y., Kim, J. P., and Rakela, J.(1997). Hepatitis G virus coinfection in hepatitis C virus- infected liver transplant recipients. *Transplantation* **64**, 786-788.

Velazquez, O., Stetler, H. C., Avila, C., Ornelas, G., Alvarez, C., Hadler, S. C., Bradley, D. W., and Sepulveda, J.(1990). Epidemic transmission of enterically-transmitted non-A, non-B hepatitis in Mexico, 1986-1987. *JAMA* **263**, 3281-3285.

Viazov, S., Riffelmann, M., Khoudyakov, Y., Fields, H., Varenholz, C., and Roggendorf, M.(1997). Genetic heterogeneity of hepatitis G virus isolates from different parts of the world. *Journal of General Virology* **78**, 577-581.

Walker, F. M., Dazza, M. C., Dauge, M. C., Boucher, O., Bedel, C., Henin, D., and Lehy, T.(1998). Detection and localization by in situ molecular biology techniques and immunohistochemistry of hepatitis C virus in livers of chronically infected patients. *Journal of Histochemistry and Cytochemistry* **46**, 653-660.

Wang, H. L., Hou, Y. D., and Jin, D. Y.(1997). Identification of a single genotype of hepatitis G virus by comparison of one complete genome from a healthy carrier with eight from patients with hepatitis. *Journal of General Virology* **78**, 3247-3253.

Wang, H. L. and Jin, D. Y.(1997). Prevalence and genotype of hepatitis G virus in Chinese professional blood donors and hepatitis patients. *Journal of Infectious Diseases* **175**, 1229-1233.

Wang, J. T., Tsai, F. C., Lee, C. Z., Chen, P. J., Sheu, J. C., Wang, T. H., and Chen, D. S.(1996). A prospective study of transfusion-transmitted GB virus C infection: Similar frequency but different clinical presentation compared with hepatitis C virus. *Blood* **88**, 1881-1886.

Wang, Y., Chen, H. S., Fan, M. H., Liu, H. L., An, P., Sawada, N., Tanaka, T., Tsuda, F., and Okamoto, H.(1997). Infection with GB virus C and hepatitis C virus in hemodialysis patients and blood donors in Beijing. *Journal of Medical Virology* **52**, 26-30.

Watson, H. G., Ludlam, C. A., Rebus, S., Zhang, L. Q., Peutherer, J. F., and Simmonds, P.(1992). Use of several second generation assays to determine the true prevalence of hepatitis C infection in haemophiliacs treated with non-virus inactivated factor VIII and IX concentrates. *British Journal of Haematology* **80**, 514-518.

Weiner, A. J., Brauer, M. J., Rosenblatt, J., Richman, K. H., Tung, J., Crawford, K., Bonino, F., Saracco, G., Choo, Q. L., Houghton, M., and Han, J. H.(1991). Variable and hypervariable domains are found in the regions of HCV

corresponding to the flavivirus envelope and NS1 proteins and the pestivirus envelope glycoproteins. *Virology* **180**, 842-848.

Westaway, E. G., Brinton, M. A., Gaidamovich, S. Y., Horzinek, M. C., Igarashi, A., Kaariainen, L., Lvov, D. K., Porterfield, J. S., Russell, P. K., and Trent, D. W.(1985). *Flaviviridae* (Report of the *Togaviridae* Study Group, Vertebrate Virus Subcommittee, International Committee on Taxonomy of Viruses). *Intervirology* **24**, 183-192.

Wu, J. C., Sheng, W. Y., Huang, Y. H., Hwang, S. J., and Lee, S. D.(1997). Prevalence and risk factor analysis of GBV-C/HGV infection in prostitutes. *Journal of Medical Virology* **52**, 83-85.

Xiang, J. H., Klinzman, D., McLinden, J., Schmidt, W. N., Labrecque, D. R., Gish, R., and Stapleton, J. T.(1998). Characterization of hepatitis G virus (GB-C virus) particles: Evidence for a nucleocapsid and expression of sequences upstream of the E1 protein. *Journal of Virology* **72**, 2738-2744.

Yanagi, M., StClaire, M., Shapiro, M., Emerson, S. U., Purcell, R. H., and Bukh, J.(1998). Transcripts of a chimeric cDNA clone of hepatitis C virus genotype 1b are infectious in vivo. *Virology* **244**, 161-172.

Yanagi, M., StClaire, M., Emerson, S. U., Purcell, R. H., and Bukh, J.(1999). *In*

vivo analysis of the 3' untranslated region of the hepatitis C virus after *in vitro* mutagenesis of an infectious cDNA clone. *Proceedings of the national Academy of Science, USA.* **96**, 2291-2295.

Yap, S. H., Willems, M., Vandenoord, J., Habets, W., Middeldorp, J. M., Hellings, J. A., Nevens, F., Moshage, H., Desmet, V., and Fevery, J.(1994). Detection of hepatitis C virus antigen by immuno- histochemical staining - a histological marker of hepatitis C virus infection. *Journal of Hepatology* **20**, 275-281.

Yoo, B. J., Selby, M. J., Choe, J., Suh, B. S., Choi, S. H., Joh, J. S., Nuovo, G. J., Lee, H. S., Houghton, M., and Han, J. H.(1995). Transfection of a differentiated human hepatoma cell line (huh7) with in vitro-transcribed hepatitis C virus (HCV) RNA and establishment of a long-term culture persistently infected with HCV. *Journal of Virology* **69**, 32-38.

Yu, H., Grassmann, C. W., and Behrens, S. E.(1999). Sequence and structural elements at the 3' terminus of bovine viral diarrhea virus genomic RNA: functional role during RNA replication. *Journal of Virology* **73**, 3638-3648.

Yuasa, T., Ishikawa, G., Manabe, S., Sekiguchi, S., Takeuchi, K., and Miyamura, T.(1991). The particle size of hepatitis C virus estimated by filtration through microporous regenerated cellulose fibre. *Journal of General Virology* **72**, 2021-2024.

Zanetti, A. R., Tanzi, E., Romano, L., Principi, N., Zuin, G., Minola, E., Zapparoli, B., Palmieri, M., Marini, A., Ghisotti, D., Friedman, P., Hunt, J., and Laffler, T.(1998). Multicenter trial on mother-to-infant transmission of GBV-C virus. *Journal of Medical Virology* **54**, 107-112.

Zhang, L. Q., Simmonds, P., Ludlam, C. A., and Leigh Brown, A. J.(1991). Detection, quantitation and sequencing of HIV-1 virus from the plasma of seropositive individuals and from factor VIII concentrates. *AIDS* **5**, 675-681.

Zignego, A. L., Macchia, D., Monti, M., Thiers, V., Mazzetti, M., Foschi, M., Maggi, E., Romagnani, S., Gentilini, P., and Brechot, C.(1992). Infection of peripheral mononuclear blood cells by hepatitis-c virus. *Journal of Hepatology* **15**, 382-386.

Zuckerman, A. J.(1996). Alphabet of hepatitis viruses. *The Lancet* **347**, 558-559.

Zuckerman, A.J. (1990). *Viral Hepatitis.* British Medical Bulletin edition. Churchill Livingstone, Edinburgh.

APPENDIX

2.1 DENSITY GRADIENT ULTRACENTRIGATION

Sucrose (AnalaR)

TEN buffer : 50mM Tris-HCl (pH 8.0)

1mM EDTA

150mM NaCl

Beckman L8-60M ultracentrifuge with SW 50.1 rotor

polyalomer ultracentrifugation tubes....(5ml)

Abbé refractometer

2.2 POLYMERASE CHAIN REACTION

2.2.1 EXTRACTION OF VIRAL RNA FROM SERUM SAMPLES

Lysis buffer: 50mM Tris-HCl (pH 8.0)

100mM NaCl

1mM EDTA

0.5% sodium-n-lauroylsarcosine

1mg/ml Proteinase K

40μg/ml polyadenylic acid

Phenol (Rathburn Chemicals Ltd)

Chloroform (AnalaR)

Isoamylacohol (BDH)

XM Sodium acetate (pH 5.2)

100% Ethanol (AnalaR)

Diethylpyrocarbonate-treated water (DEPC-water)

Water-bath at 37^0C

Centrifuge (Hereus Biofuge 15R)

2.2.1 EXTRACTION OF RNA FROM CULTURED CELLS

Denaturing Solution (Promega): 26mM sodium citrate (pH 4.0)

0.125M β-mercaptoethanol

0.5% N-lauryl sarcosine

4M guanidine thiocyanate

Phenol:Chloroform:Isoamylalcohol (125:24:1) buffered with 42mM sodium citrate (pH 4.7) (Promega)

Isopropanol

2M Sodium acetate (pH 4.0)

Nuclease-free water

2.2.3 REVERSE-TRANSCRIPTION OF VIRAL RNA

RT buffer: 50mM Tris-HCl (pH 8.0)

5mM $MgCl_2$

5mM ditiothreithol

50mM KCl

0.05μg/μl bovine serum albumin

15% dimethylsulfoxide

Nucleoside triphosphate mixture (100mM stock; dilute to 4mM; Boehringer Mannheim)

Rnasin (Rnase inhibitor; Promega-0.5μl is equal to 10 units)

Avian Myeloblastosis Virus Reverse Transcriptase (AMV; Promega-1μl is equal to 10 units)

Liquid paraffin (William Ranson&son)

Thermal cycler Techne GeneE

2.2.4& 2.2.5 PCR AMPLIFICATION

10 X PCR reaction buffer: 100mM Tris-HCl (pH 9.0)

500mM KCl

15mM $MgCl_2$

0.5% Triton X-100

Nucleoside triphosphate mixture (100mM stock; dilute to 3mM)

Taq polymerase (Promega; 1μl is equal to 5 units)

Liquid paraffin

Thermal cycler (Techne GeneE)

2.2.6 ANALYSIS OF PCR PRODUCTS

2% agarose gel: 6g agarose (Flowgen)

300ml 1 x TBE

10μl Ethidium Bromide (Sigma)

10 x TBE: 108g Tris base (AnalaR)

55g Boric acid (Molecular Biology Certified; Kodak)

40ml 0.5M EDTA (Molecular Biology Certified; Kodak)

Make up to 1L with distilled water

1 x TBE used as electrophoresis buffer in the gel tank.

2.3 DNA SEQUENCING

2.3.1 PREPARATION OF DNA FOR SEQUENCING

2.3.1.1 SOLID PHASE PURIFICATION OF PCR PRODUCTS

Binding and washing buffer (BW): 10mM Tris-HCl (pH 7.5)

1mM EDTA

2M NaCl (final concentration 1M)

0.15M NaOH

TE buffer: 10mM Tris-HCl (pH 7.5)

1mM EDTA

Dynabeads M-280 streptavidin (Dynal)

Dynal magnetic particle concentrator 6 (MPC-6; Dynal)

2.3.1.2 ALKALINE DENATURATION OF PLASMID DNA

Denaturation solution: 2M NaOH

2mM EDTA

3M Sodium acetate (pH 4.5-5.5)

100% Ethanol (AnalaR)

DEPC-water

2.3.2 DIRECT SEQUENCING OF AMPLIFIED PCR PRODUCTS

5 x Sequenase reaction buffer: 200mM Tris-HCl (pH 7.5)

100mM $MgCl_2$

250mM NaCl

5 x Labelling mix: 7.5μM dGTP

7.5μM dTTP

7.5μM dCTP

[α^{35}S] dATP (Amersham)

0.1M DTT

Termination mixtures : 80μM of all four dNTPs supplemented with 8μM ddATP, ddTTP, ddGTP or ddCTP.

Stop solution: 95%formamide

20mM EDTA

0.05% xylene cyanol

0.05% bromophenol blue

Sequenase Version 2.0 T7 DNA Polymerase (1μl is equal to 13 units) in 20mM KPO_4 (pH 7.4), 1mM DTT, 0.1mM EDTA, 50%Glycerol.

2.3.4 DENATURING PAGE

5% Denaturing PAGE gel: 21g urea (AnalaR)

6ml Ultrapure Sequagel concentrate (50%; National Diagnostics)

5ml 10 x TBE (Sanger)

0.05g Ammonium persulfate (APS; Sigma)

Make up to 50ml with dH_2O

Add 25μl TEMED (Sigma) prior to puring gel mix.

10 x TBE (Sanger; 2litre): 324g Tris base

85g Boric acid

19g EDTA

Make up to 2l with dH_2O.

BIOMAX autoradiography film (Eastman Kodak).

2.4 CLONING OF PCR PRODUCTS

Geneclean kit (BIO 101 Inc.): 6M Sodium iodide

New Concentrate (NaCl/Tris/EDTA)

GLASSMILK suspension (silica matrix in water)

100% Ethanol

TE buffer

10 x Ligase buffer: 200mM Tris-HCl (pH 7.6)

50mM $MgCl_2$

10mM ATP

T4 DNA ligase (0.5μl is equal to 2-3 Weiss units; R&D SystemsLtd.)

Competent cells derived from DHI cells

Luria Broth (LB) medium: 10g Bacto-tryptone (DIFCO)

5g Bacto-yeast extract (GIBCO.BRL)

10g NaCl

Dissolve in H_2O and adjust to pH 7.0 using NaOH. Make up 1 liter with H_2O. Sterilize by autoclaving for 20 minutes at 15psi on a liquid cycle.

LB agar: Add 15g agar (Bacteriological grade;ICN) to 1 liter of LB medium and sterilize by autoclaving for 20 minutes. Allow to cool to 40-50^{0}C before adding the antibiotics and IPTG/X-Gal.

100mM IPTG stock solution: 24mg IPTG per ml H_2O

filter sterilize and keep on ice until ready to use.

5% X-Gal stock solution: 50mg X-Gal per ml of N,N' dimethyl-formamide in a sterile glass bottle. Cover with silver foil and store on ice until ready to use.

Solution GTE (100ml): 0.9g glucose

2ml 0.5M EDTA

2.5ml 1M Tris HCl

NaOH/SDS solution (10ml): 400 μl 5M NaOH

8.6ml distilled water

1ml 10% SDS

Solution III: 29.5ml glacial acetic acid (AnalaR-BDH)

adjust to pH 4.8 with KOH pellets (Analar-BDH)

made up to 100ml with distilled water.

2.5 CELL CULTURE

RPMI 1640 growth medium (Gibco BRL)

Supplements for 100ml medium: 10% FCS (Gibco BRL)

100mM L-Glutamine (Gibco BRL)

10^4IU/ml penicillin/streptomycin (Gibco BRL)

10^4IU/ml interleukin-2 (NIBSC)

5μg/ml phytohaemagglutinin (Murex)

phorbol 12-myristate, 13-acetate (Sigma) at a final concentration of 160nM

Ham's F-12 minimal medium (Life Technologies ltd.)

Supplements for 100ml medium: 50ng/ml epidermal growth factor (ICN Pharmaceuticals ltd.)

1ng/ml glucagon (ICN Pharmaceuticals ltd)

10μg/ml insulin (Life Technologies ltd)

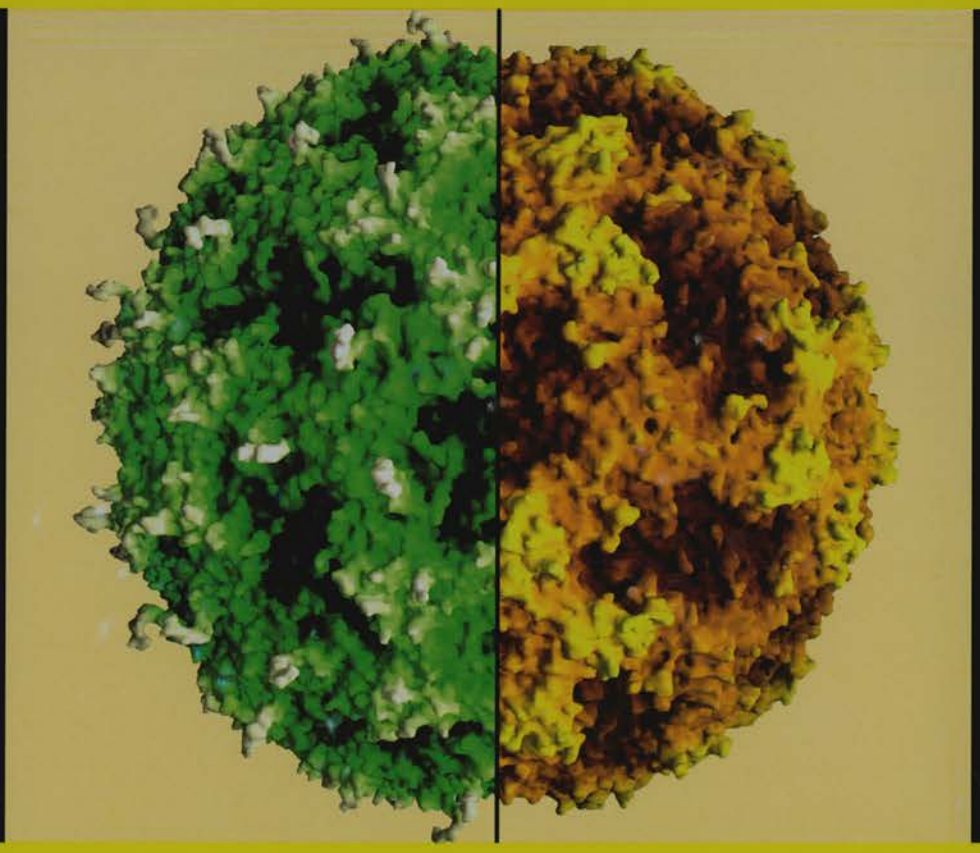
6.5ng/ml somatostatin(Sigma)

3.5μM hydrocotisone (ICN Pharmaceuticals ltd.)

5μg/ml transferrin (Life Technologies ltd.)

Journal of General Virology

Volume 78 1997



1967–1997: Thirty years of excellence

Offprint

Journal of General Virology (1997), **78**, 1533–1542. Printed in Great Britain

Discrimination of hepatitis G virus/GBV-C geographical variants by analysis of the 5′ non-coding region

Donald B. Smith,[1] Narcisa Cuceanu,[1] Fiona Davidson,[2] Lisa M. Jarvis,[1] John L. K. Mokili,[1] Saeed Hamid,[3] Christopher A. Ludlam[4] and Peter Simmonds[1]

[1] Department of Medical Microbiology, University of Edinburgh, Medical School, Teviot Place, Edinburgh EH8 9AG, UK
[2] Edinburgh and South East Scotland Blood Transfusion Service, Lauriston Place, Edinburgh EH3 9HB, UK
[3] Department of Medicine, The Aga Khan University Hospital, Stadium Road, Karachi 74800, Pakistan
[4] Department of Haematology, Royal Infirmary of Edinburgh, Lauriston Place, Edinburgh EH3 9YW, UK

We have investigated the ability of different subgenomic fragments to reproduce the phylogenetic relationships observed between six complete genome sequences of GBV-C/hepatitis G virus (HGV). While similar relationships were observed following analysis of part of the 5′ non-coding region (5′NCR), for the coding region they were not accurately reproduced for some large fragments or for the majority of fragments of 300 or 600 nucleotides. Analysis of 5′NCR sequences from a large number of isolates, including newly obtained sequences from Pakistan, Zaïre and Scotland, produced separate groupings of Asian, African and European/North American variants. These groupings are associated with specific polymorphisms in the 5′NCR, many of which were covariant and consistent with a proposed secondary structure for this region. The relatively low level of amino acid sequence variation observed between these geographically and phylogenetically defined groups of variants suggests that they are unlikely to display significant biological differences.

Introduction

Investigation of the genetic heterogeneity of GBV-C/hepatitis G virus (HGV) (Simons *et al.*, 1995; Linnen *et al.*, 1996) is at an early stage. This newly discovered human virus has a genome structure related to that of hepatitis C virus (HCV) with a 5′ non-coding region (5′NCR) capable of acting as an internal ribosome entry site (IRES) (Simons *et al.*, 1996) followed by a long open reading frame capable of encoding presumed structural (E1 and E2) and nonstructural (NS2, NS3, NS4a, NS4b, NS5a and NS5b) proteins (Simons *et al.*, 1995; Linnen *et al.*, 1996). GBV-C/HGV RNA can be detected in 2–4% of blood donors throughout the world (Jarvis *et al.*, 1996; Linnen *et al.*, 1996; Moaven *et al.*, 1996; Schlueter *et al.*, 1996; Stark *et al.*, 1996), but an association with disease in infected individuals has yet to be established.

Analysis of the 5′NCR has led to the suggestion that variants of GBV-C/HGV can be divided into three or more 'genotypes' (Fukushi *et al.*, 1996; Muerhoff *et al.*, 1996). Some of these groupings are associated with the geographical origin of the variant, but otherwise nothing is known about the virological or clinical implications of genomic differences between them. Attempts to demonstrate phylogenetic groupings of GBV-C/HGV isolates based on sequence analysis of subgenomic coding regions have generally been unsuccessful. For example, comparison of sequences within the NS3 helicase region for isolates from around the world failed to differentiate between Asian, African and European/North American isolates (Tsuda *et al.*, 1996; Pickering *et al.*, 1997), and as much diversity was observed within isolates from Taiwan as between isolates from different continents (Kao *et al.*, 1996). Similarly, analysis of a fragment of NS5b failed to distinguish between African and Asian isolates, and as much variation was observed between variants from Russia or Germany as between isolates from different continents (Viazov *et al.*, 1997). In neither case was there a difference in the distribution of sequence distances within groups compared to between groups. In contrast, for HCV three non-overlapping distributions corresponding to virus type, subtype and isolate have been described for coding regions (Simmonds *et al.*, 1993*a*, 1994), but these groupings are less clear from analysis

Author for correspondence: Donald Smith.
Fax +44 131 650 6531. e-mail Donald.B.Smith@ed.ac.uk

The GenBank accession numbers of the sequences reported in this paper are AF003152–AF003186.

of the 5′NCR (Simmonds *et al.*, 1993*b*; Ohba *et al.*, 1995), probably because of its extreme conservation and the existence of covariant substitutions (Smith *et al.*, 1995).

In this study we have used the complete genome sequences of six different GBV-C/HGV isolates to investigate the ability of different subgenomic regions to reproduce the phylogenetic relationships displayed by comparison of the complete sequences. While these relationships are not reproduced by analysis of short fragments of coding regions, they can be reproduced by comparison of 5′NCR sequences. We describe the features of variability within the 5′NCR that may be useful in the development of methods to discriminate between different variant groupings.

Methods

■ **Samples.** Sera were obtained from 14 patients infected with GBV-C/HGV with chronic hepatitis from Pakistan, 12 of whom were co-infected with HCV or hepatitis B virus, and from nine women from Zaïre who were co-infected with human immunodeficiency virus. Samples of plasma were also available from 12 haemophiliacs previously treated in Edinburgh with locally manufactured clotting factor concentrates not inactivated for enveloped viruses (Jarvis *et al.*, 1996). Two haemophiliacs had received commercially produced concentrates, one exclusively (Ed 3), while the other (Ed 81) had also received locally produced, but inactivated concentrates. Samples were stored at −70 °C before extraction of RNA using proteinase K–Sarkosyl and phenol–chloroform extraction as described previously (Jarvis *et al.*, 1994).

■ **RT–PCR and sequencing.** Purified RNA was reverse-transcribed and amplified by nested RT–PCR using primers derived from the 5′NCR of GBV-C/HGV: S4571 – sense outer (positions −445 to −428, see below for method of numbering), S4572 – sense inner (positions −419 to −399), S4573 – antisense inner (positions −76 to −97) and S4574 – antisense outer (positions −22 to −42) (Jarvis *et al.*, 1996). RNA from the haemophiliac samples was also amplified with primers T1721 – sense outer (5′ GGGCAAACGACGCCCACGTACGGTC 3′, positions −244 to −220), T1722 – sense inner (5′ TCGCCCTTCAATGYCTCTC-TTGRCC 3′, positions −215 to −191), T1723 – antisense inner (5′ GTGCACCCCAGRGCCACMAGGCA 3′, positions 142 to 164) and T1724 – antisense outer (5′ CCCGCCTGATACAGYGGCCAGCA 3′, positions 181 to 203).

Reverse transcription was performed using avian myeloblastosis virus reverse transcriptase (Promega) at 42 °C for 30 min, while conditions for PCR were hot start at 80 °C for 2·5 min followed by 30 cycles of 94 °C for 18 s, 55 °C for 21 s and 72 °C for 90 s. Secondary PCR was carried out using 1 μl of the primary PCR product, and amplified products were visualized after electrophoresis through a 2% agarose gel stained with ethidium bromide.

■ **Nucleotide sequences.** PCR products were re-amplified from primary products using primers S4572 and biotinylated S4573 and sequenced directly from magnetically separated single strands after immobilization on streptavidin-coated beads (Dynabeads, Dynal). Secondary PCR products produced using primers T1722 and T1723 were purified and cloned using the LigATor kit (R&D Systems), and sequenced using T7 DNA polymerase (Sequenase, USB) and both sense and antisense plasmid primers. The consensus sequence of one to five clones was used for phylogenetic analysis. Other nucleotide sequences obtained from GenBank are identified by their accession number, or by isolate name for the six complete genome sequences [GBV-C, U36380; PNF2161, U44402; R10291, U45966; GBV-C(EA), U63715; HGVC964, U75356; and HGV-Iw, D87255].

■ **Numbering of nucleotide positions.** The true 5′ terminus of the GBV-C/HGV genome is currently unknown. Although the sequence of PNF2161 extends the furthest upstream, the existence of additional 5′ sequences is suggested by the lack of a well-defined 5′-terminal hairpin structure similar to those defined for GBV-A (Simons *et al.*, 1996) and for different genotypes of HCV (Smith *et al.*, 1995). Because of this uncertainty we have adopted a system similar to that commonly used for describing nucleotide positions within the HCV genome. Positions are given relative to the AUG codon near the start of the long open reading frame which follows the multiple stem–loop structures and the polypyrimidine tract of the internal ribosome entry site (IRES) (Simons *et al.*, 1996). This AUG (positions 524–526 of Simons *et al.*, 1996) is likely to represent the initiation codon for translation of the GBV-C /HGV polyprotein because of its position relative to the IRES, the initiation of *in vitro* translation products at this site (Simons *et al.*, 1996), and the lack of a conserved upstream open reading frame or of an alternative in-frame AUG codon (Muerhoff *et al.*, 1996). Positions in the coding region are numbered from the AUG of the prototype sequence (GBV-C) while positions in the 5′NCR are given negative numbers relative to this AUG. The 5′NCR contains 11 sites of insertion relative to the GBV-C sequence in our data set, and these positions are un-numbered.

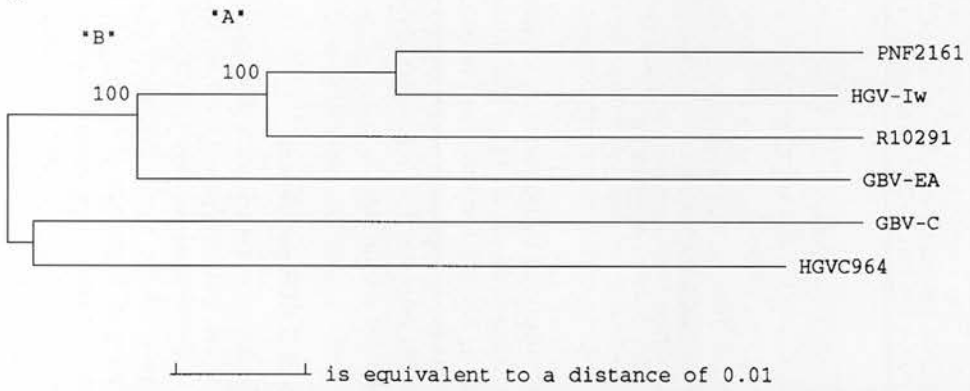
■ **Sequence analysis.** Sequences were aligned using Esee (version 1.09, Eric Cabot) or Simmonic Software (version 1.0, Peter Simmonds). Phylogenetic trees were produced using the MEGA package (Kumar *et al.*, 1993). Free energies for stem–loop structures were obtained using the program FOLDRNA in the GCG package (Genetics Computer Group, Wisconsin, USA, accessed via the SEQNET facility of Daresbury).

Results

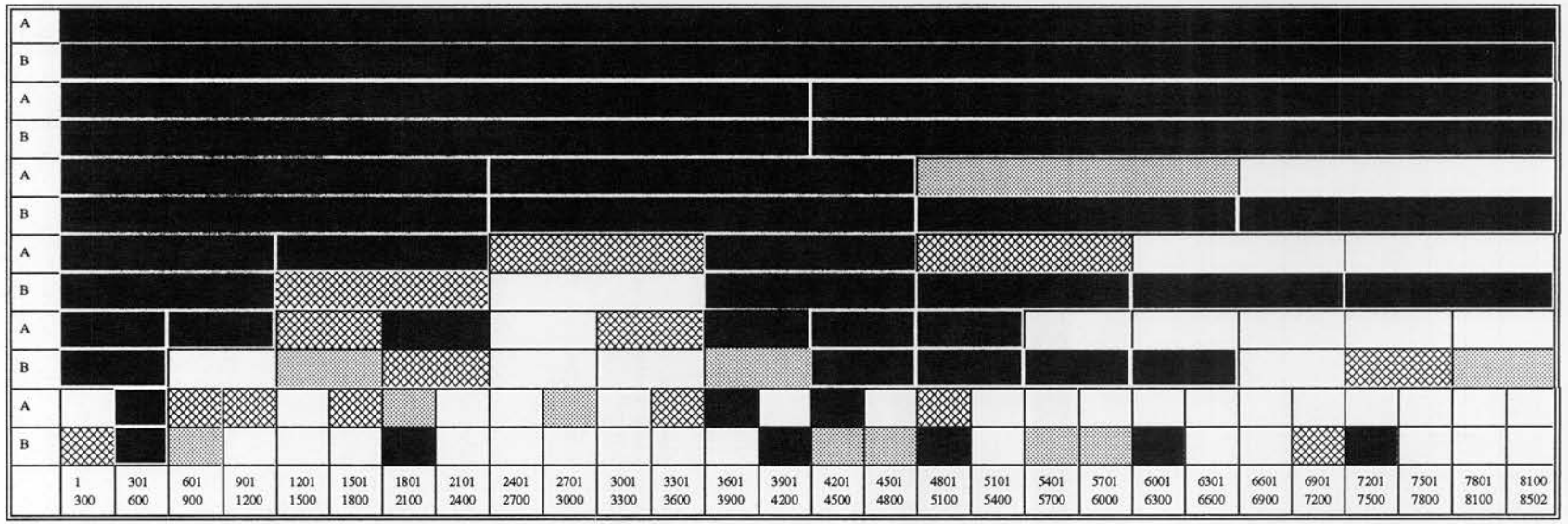
Phylogenetic analysis of complete coding sequences

Three major 'types' of GBV-C/HGV have been described based on analysis of the 5′NCR, two of which were further divided into 'subtypes' (Muerhoff *et al.*, 1996). This analysis included three variants for which complete genome sequences are available, namely GBV-C ('type 1'), and PNF2161 and R10291 ('subtype 2a'). Three additional complete genome sequences are now available, and we investigated their phylogenetic relationships through analysis of their coding regions (Fig. 1*a*). HGV-Iw, isolated from a Japanese hepatitis patient (Shao *et al.*, 1996), is more closely related to PNF2161 than to R10291, and grouped with these sequences in 100% of bootstrap re-sampling replications, and so is 'subtype 2a'. GBV-C(EA), isolated from an East African child, is more divergent but grouped with the 'subtype 2a' variants in 100% of bootstrap replications, and shares a higher degree of sequence identity with these sequences than with GBV-C (Erker *et al.*, 1996). Comparison of the Chinese variant HGVC964 with the other five complete genome sequences revealed two areas where numerous nucleotide and amino acid substitutions occur (positions 5917–5948, and 8502 onwards). The strong conservation of five out of the six sequences available for these regions suggests that sequencing errors could be responsible for the extreme variation of HGVC964 in these regions. A perfect alignment of the amino acid sequence

(a)



(b)



Bootstrap level (%)	91-100	81-90	71-80	< 71

Fig. 1. Phylogenetic analysis of six complete GBV-C/HGV genomes. (*a*) Phylogenetic tree of the coding region (positions 1–8502) of six complete GBV-C/HGV genomes. The percentage of bootstrap re-sampling replicates ($n = 5000$) in which branches 'A' and 'B' were observed is indicated. (*b*) Bootstrap support for phylogenetic groupings based on analysis of subgenomic fragments from the coding region of complete genome sequences. Phylogenetic analysis was performed on the complete coding region and on progressively smaller subgenomic fragments from 4200 to 300 nt. The percentage of bootstrap replicates ($n = 500$) in which PNF2161, R10291 and HGV-Iw shared a common branch (A) or in which these sequences and GBV-C(EA) shared a common branch (B) is indicated by different patterns. Bootstrap frequencies of less than 70% are regarded as not providing evidence for the phylogenetic grouping. Nucleotide positions are numbered relative to the AUG initiation codon of GBV-C.

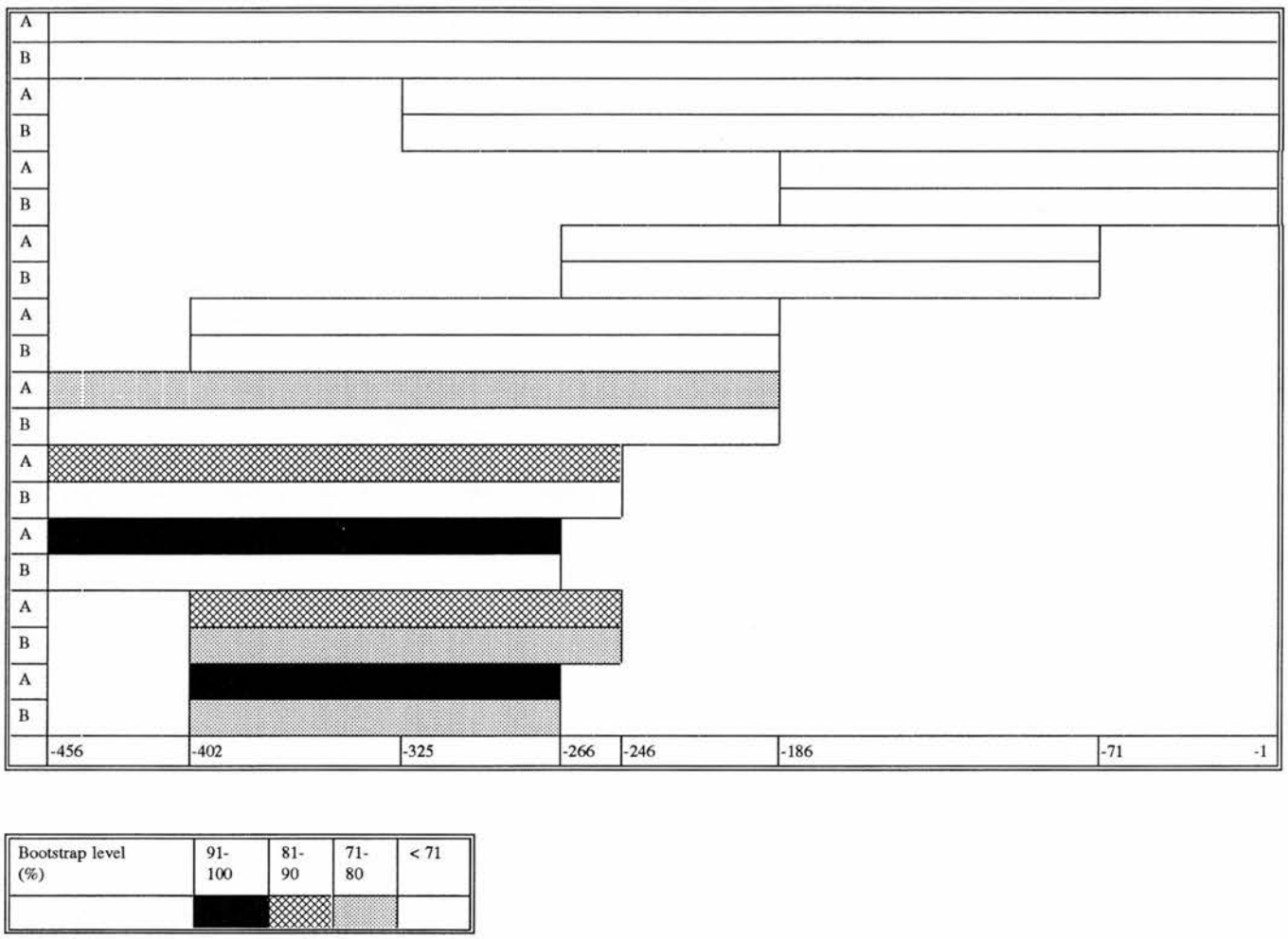


Fig. 2. Level of bootstrap support for phylogenetic groupings based on analysis of fragments of the 5′NCR. The level of bootstrap support (500 replications) for the grouping of PNF2161, R10291 and HGV-Iw (A) or of these three sequences and that of GBV-C(EA) (B) is indicated by different patterns. The 5′ and 3′ boundaries of the regions analysed are indicated at the bottom of the figure, numbered relative to the initiator AUG of GBV-C.

encoded by positions 5917–5948 can be obtained by shifting the HGVC964 sequence by one nucleotide. Comparisons between the complete genome sequences have therefore been made by including this frameshift and by removing the second discrepant region for which no obvious realignment could be found. These alterations had little effect on the topography of the trees generated and HGVC964 did not group with any of the other sequences. We present evidence below from analysis of 5′NCR sequences that GBV-C(EA) belongs to 'subtype 2b' and that HGVC964 represents a new group of variants distinct from the Japanese variants which form 'type 3'.

We next investigated the extent to which analysis of different subgenomic fragments supported the grouping of isolates PNF2161, R10291 and GBV-Iw (A), or the grouping of these isolates and GBV-C(EA) (B) in bootstrap re-sampling replications (Fig. 1*b*). Two 1800 nucleotide (nt) fragments in the 3′-half of the coding region failed to support the grouping of PNF2161, R10291 and GBV-Iw sequences, while one or both groupings were not supported by bootstrap re-sampling in 3/6 (50%) 1200 nt fragments, in 8/14 (57%) 600 nt fragments or in 23/28 (82%) 300 nt fragments. The five 300 nt regions that did provide bootstrap support for both groupings were positions 301–600 (E1), 601–900 (E2), 1801–2100 (NS2), 4201–4500 (NS3) and 4801–5100 (NS4a).

Phylogenetic analysis of 5′NCR sequences

Because of the difficulties associated with phylogenetic analysis of coding regions, and since analysis of the 5′NCR has previously been shown to discriminate between GBV-C/HGV isolates from different parts of the world (Muerhoff *et al.*, 1996), we tested the ability of fragments of the 5′NCR to reproduce the phylogenetic relationships observed amongst complete coding sequences (Fig. 2). Comparison of the largest region for which sequences were available for all six sequences (positions −456 to −1) failed to provide support for the

grouping of PNF2161 and HGV-Iw with R10291 (A), or for the grouping of these sequences with GBV-C(EA) (B). These groupings were also not observed for fragments between −325 and −1, but were supported by analysis of 5′NCR fragments between positions −402 to −246. This represents less than a third of the 5′NCR, and is much smaller than that of coding regions giving similar degrees of bootstrap support (from 300 to 4200 nt depending on the region).

Geographical distribution of GBV-C/HGV groupings

In order to test the ability of phylogenetic analysis of a restricted portions of the 5′NCR to discriminate between GBV-C/HGV isolates from different geographical regions, we analysed sequences available from GenBank and new sequences obtained from pregnant women from Zaïre ($n = 9$), patients with chronic hepatitis from Pakistan ($n = 14$) and haemophiliacs treated in Edinburgh ($n = 12$). Separate clusters of African, Asian and European/North American sequences were observed consistently when 5′-terminal regions of the 5′NCR were analysed. For the region −366 to −235, virus sequences from Zaïre were similar to those present in published sequences from West Africa, while those from Pakistan and from most of the Edinburgh haemophiliacs grouped with sequences previously described from infected individuals in Europe and North America (Fig. 3). The only exceptions were two haemophiliacs infected with variants related to the African (Ed 3) or Asian groupings (Ed 81), both of whom had received commercial factor concentrates, and in contrast to the other haemophiliacs who had only received factor concentrates manufactured from blood donations collected locally. Two other exceptions from the general pattern were GBV-C(EA) from a child in East Africa with acute non-A–E hepatitis, and HGV-Iw from a Japanese hepatitis patient. Complete genome sequences are available for both of these isolates, and phylogenetic analysis of the full-length coding sequences (Fig. 1*a*) is consistent with the groupings obtained by analysis of 5′NCR sequences. Phylogenetic analysis of fragments of a similar size from the 3′ terminus of the 5′NCR did not consistently produce the same groupings of isolates as those based on geographical origin (data not shown). These observations suggest that analysis of the region −366 to −235 accurately reproduces the phylogenetic relationship of GBV-C/HGV isolates, consistent with the analysis of similar regions from the complete genome sequences (Fig. 2).

5′NCR polymorphisms correlated with different phylogenetic groupings

We next searched for polymorphisms in the 5′NCR that were associated with these phylogenetic groupings since they might simplify the identification of GBV-C/HGV groups from sequence information, or allow the development of methods of identification that avoid the need for sequence analysis. Polymorphisms that are strongly correlated with phylogenetic groupings are clustered in three separate regions of the 5′NCR (Fig. 4), and the majority of these were consistent with a proposed secondary structure for GBV-C/HGV 5′NCR (Simons *et al.*, 1996) since substitutions either maintained proposed base pairings, or occurred in unpaired regions. Group-specific polymorphisms are also present outside these regions (positions −65, −103, −104, −108, −149, −306, −307), but the remaining polymorphic positions vary without regard to group. For example, substitutions within the unpaired polypyrimidine stretch between positions −21 to −10 generally are confined to alteration between U or C residues and no group-specific pattern can be discerned. Another such region is that between positions −147 to −120, where despite the considerable sequence variability, all sequences are capable of forming stem–loop structures with free energies of between −42·6 and −89·0 kJ (Fig. 5). Some substitutions in this region are group-specific, but at the majority of polymorphic sites, substitutions occur in all of the phylogenetic groups. Similarly, most substitutions between positions −185 and −177 do not correlate with the phylogenetic groupings based on analysis of complete genome sequences, but the ability to form a short stem–loop structure (region IVa′ of Simons *et al.*, 1996) is conserved.

Discussion

Variation of GBV-C/HGV coding regions

Our phylogenetic analysis of six complete genome sequences of GBV-C/HGV has revealed several unexpected findings. While certain phylogenetic groupings are apparent from the analysis of the complete coding region, these groupings are only poorly supported when smaller genome segments are compared. These observations provide an explanation for the previous inability to observe consistent phylogenetic relationships between GBV-C/HGV isolates from different parts of the world based on analysis of 118 nt within NS3 (positions 3767–3884) (Kao *et al.*, 1996; Pickering *et al.*, 1997) or from the analysis of 354 nt within NS5b (positions 6742–7095) (Viazov *et al.*, 1997). Analysis of similar genome regions from complete sequences revealed that neither a 300 nt fragment of NS3 nor a 1200 nt fragment of NS5 were able to reproduce the phylogenetic relationships of the complete coding sequence. This finding suggests that the use of short coding sequences such as 118 nt or less of NS3 in investigations of the epidemiology of GBV-C/HGV (Berg *et al.*, 1996; Heringlake *et al.*, 1996; Masuko *et al.*, 1996; Schmidt *et al.*, 1996; Schreier *et al.*, 1996; Tsuda *et al.*, 1996) may not be reliable for the reconstruction of phylogenetic relationships. The finding that analysis of relatively long coding regions is required to distinguish between different groups of GBV-C/HGV variants differs markedly from the situation for HCV, for which phylogenetic groupings observed by analysis of complete virus genomes are closely mirrored by those observed for subgenomic fragments throughout the genome,

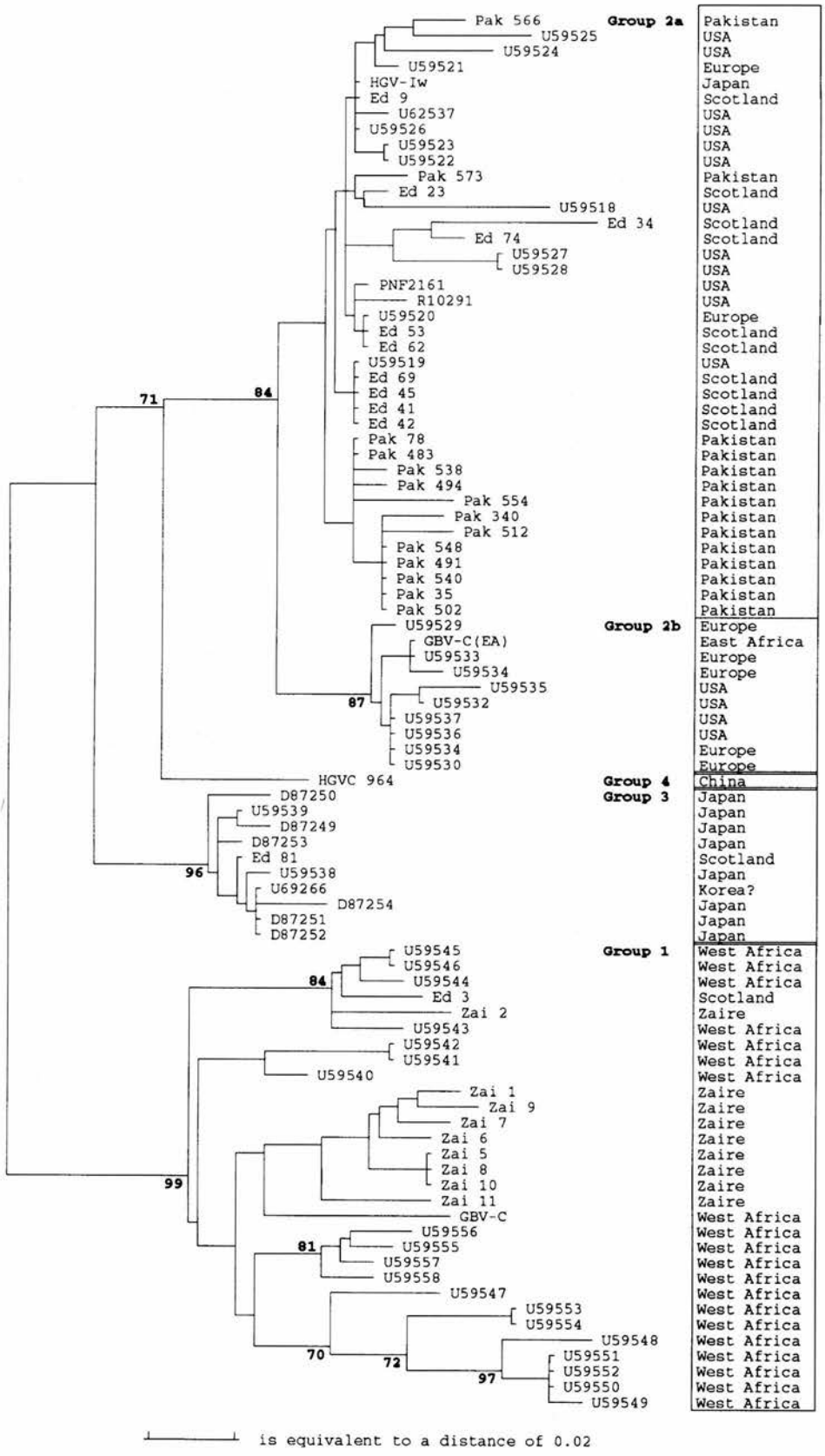


Fig. 3. For legend see facing page.

Fig. 4. Group-specific 5′NCR polymorphisms. Sequences for the regions −490 to −459, −366 to −329 and −287 to −235 are grouped on the basis of phylogenetic analysis of positions −366 to −235. Sequence identity with PNF2161 is indicated by dots, missing information by spaces and ambiguous or mixed nucleotides by question marks.

Fig. 3. Consensus phylogenetic tree of geographically disparate isolates. The tree was generated by comparison of positions −366 to −235 of the 5′NCR. The percentage of bootstrap replicates in which major groupings were observed amongst 500 replicates is indicated. Sequences are identified by accession numbers or, for the sequences reported in this study, by the prefixes Pak for Pakistan, Ed for Edinburgh and Zai for Zaïre. The geographical origin of each sequence is indicated at the right.

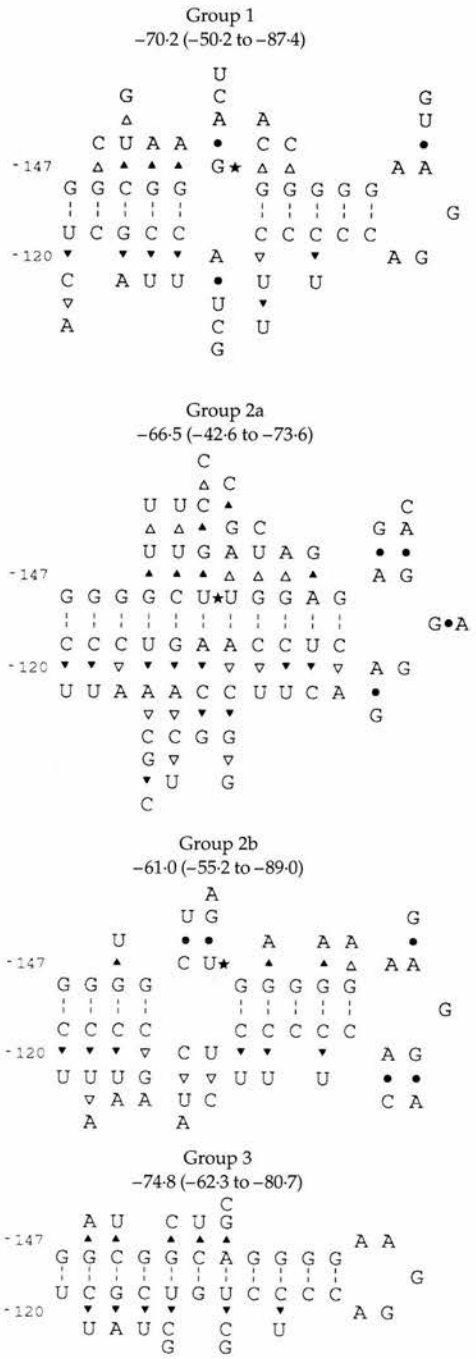


Fig. 5. Predicted stem–loop structures for different GBV-C/HGV groups. The stem–loop structure for positions −147 to −120 (corresponding to region IVb in the proposed structure of Simons *et al.*, 1996) is shown for the consensus sequence of each sequence group. Substitutions that do not affect proposed base pairings, or which are covariant with substitutions that preserve the base pairings are indicated by (▲), while substitutions that disrupt base pairings are indicated by (△). Substitutions that occur in unpaired regions are indicated by (●) while positions of nucleotide insertion relative to the consensus sequence are indicated by (★). The single Group 4 sequence is included with Group 2a. The free energy (kJ) of the stem–loop for the consensus sequence of each group is indicated, with the range observed amongst variants within each group given in parentheses.

and as small as 222 nt (Simmonds *et al.*, 1994, 1996; Ohba *et al.*, 1995).

Variation of the GBV-C/HGV 5′NCR

The extent of variation is unequal across the 5′NCR, with some areas being relatively well-conserved and others being more variable, as previously noted (Erker *et al.*, 1996; Muerhoff *et al.*, 1996). However, in contrast to the lack of discrimination of short fragments of coding regions, analysis of 5′-terminal fragments of the 5′NCR can be used to reproduce the phylogenetic groupings displayed by complete genome segments (Fig. 2). This conclusion is also supported by analysis of a large set of sequences ($n = 89$), since isolates generally grouped according to their geographical origin (Fig. 3). On the basis of this analysis the following groups can be distinguished: Group 1, a diverse collection of African sequences, Group 3 consisting of relatively similar Asian sequences, Group 2 consisting of sequences from North America, Europe and the Indian subcontinent, as well as HGV-Iw from Japan and GBV-C(EA) from East Africa. This last isolate together with some other European and North American isolates may represent a subgroup (Group 2b) since they share an unusual pattern of substitution in the 5′NCR (Fig. 4). Similarly, there is evidence for subgroupings amongst the diverse sequences comprising Group 1. A single Chinese sequence (GBV-C 964) may represent a distinct group (referred to here as group 4) since analysis of the complete coding sequence confirms that it is distinct from Groups 1 and 2. Although comparison with the complete coding sequences of Group 3 isolates is not yet possible, this sequence contains several unusual substitutions in the 5′NCR (Fig. 4).

The finding that 5′NCR polymorphisms are associated with phylogenetic groupings obtained by analysis of complete coding regions and that group according to geographical origin makes it possible to develop rapid methods for identifying GBV-C/HGV isolates, similar to those developed for the genotyping of HCV (Stuyver *et al.*, 1993; McOmish *et al.*, 1994). Group-specific polymorphisms are present throughout the 5′NCR, but they are most frequent between positions −490 and −235 ($n = 39$) where they are concentrated into three clusters (Fig. 4). Sequence information for the region between positions −490 and −459 may be more difficult to obtain because very little upstream sequence information is currently available from which to design universal primers for RT–PCR amplification. Although there are nine group-specific polymorphisms between positions −234 and −1, this region also contains three highly polymorphic regions that vary without regard to geographical grouping, and so phylogenetic analysis of this region fails to reproduce the relationships of complete coding sequences (Fig. 2) or to produce geographical groupings (data not shown). Similarly, a previous study failed to observe an association between phylogenetic groupings and geographical origin for the region −267 to +17

(Pickering *et al.*, 1997). Consequently, the optimal region for identifying the group of GBV-C/HGV isolates at present appears to be the region between positions −490 and −235, encompassing three blocks of group-specific polymorphisms.

Significance of phylogenetic groupings

The existence of distinct groups of GBV-C/HGV variants is implied by phylogenetic analysis of complete genome sequences, and by the presence of polymorphisms in the 5′NCR that correlate with geographical origin. However, the biological significance of these groupings is presently unclear, especially since representatives of different GBV-C/HGV Groups 1, 2 and 4 differ over their complete coding regions at only 3·7–4% of amino acid positions. This level of amino acid sequence variation is much less than that observed for individual genes between different serotypes of vesicular stomatitis virus (30–50%), poliovirus (20–30%) or dengue virus (23–38%), or between different types of papillomaviruses (10–17%) or subtypes of HIV (12–36%). Relative to HCV, variation is at the lower end of the range observed between complete polyprotein sequences of epidemiologically unrelated isolates of the same HCV subtype (3·5–8·1%), and much less than that observed between different HCV subtypes (13–19%) or between different HCV types (23–29%). While different types of HCV show substantial variation in antigenicity and differ in clinical features such as the frequency with which infection can be eliminated by interferon treatment, similar differences have not been associated with the more minor differences between HCV subtypes such as subtypes 1a and 1b, or the still more subtle differences between isolates of the same subtype. Comparison with HCV and other viruses would therefore suggest that the limited diversity between GBV-C/HGV variants is unlikely to be associated with major biological or clinical differences. For this reason we have used the descriptive terms 'group' and 'subgroup' rather than the terms 'type' or 'subtype' (Berg *et al.*, 1996; Fukushi *et al.*, 1996; Muerhoff *et al.*, 1996; Schreier *et al.*, 1996) that suggest that variation between phylogenetic groups is significant.

While there is relative conservation of the GBV-C/HGV polyprotein between different groups of variants, their sequences still vary by 11–13% overall because of substitution at synonymous sites. The ratio of nonsynonymous to synonymous evolutionary distances between different variant groups ranges from 0·029 to 0·054 (mean 0·04), compared with between 0·1 and 0·14 for different representatives of HCV subtype 1a. The reason for this difference between GBV-C/HGV and HCV is not known, but presumably reflects differences in selective pressures related to their interaction with the host immune system during infection. Evolutionary distances at synonymous sites between variants of GBV-C/HGV (0·56–0·7) are greater than those observed between different isolates of HCV subtype 1a (0·14–0·17) or subtype 1b (0·24–0·36) but less than those observed between different subtypes (1·02–1·65). For HCV, evolutionary distances at synonymous sites have been used to date the time of divergence of different subtypes at more than 300 years ago (Smith *et al.*, 1997). A similar estimate is not possible for GBV-C/HGV since the rate of evolution is currently unknown, but the substantial synonymous distances between sequences from different variant groups as well as their strong correlation with geographical origin, suggest that divergence of these groups is not a recent phenomenon.

We are grateful to Sue Graham for technical assistance, to staff at the Royal Infirmary of Edinburgh Haemophilia Centre for collecting samples, to Sergei Viazov for providing data before publication, and to Scott Muerhoff for clarifying the origin of certain variants. This work was supported by a grant from the Wellcome Trust to D.B.S. and P.S., and by a grant to N.C. from the Darwin Trust. P.S. is a Darwin Research Fellow.

References

Berg, T., Dirla, U., Naumann, U., Heuft, H. G., Kuther, S., Lobeck, H., Schreier, E. & Hopf, U. (1996). Responsiveness to interferon alpha treatment in patients with chronic hepatitis C coinfected with hepatitis G virus. *Journal of Hepatology* **25**, 763–768.

Erker, J. C., Simons, J. N., Muerhoff, A. S., Leary, T. P., Chalmers, M. L., Desai, S. M. & Mushahwar, I. K. (1996). Molecular cloning and characterization of a GB virus C isolate from a patient with non-A–E hepatitis. *Journal of General Virology* **77**, 2713–2720.

Fukushi, S., Kurihara, C., Ishiyama, N., Okamura, H., Hoshino, F. B., Oya, A. & Katayama, K. (1996). Nucleotide sequence of the 5′ noncoding region of hepatitis G virus isolated from Japanese patients: comparison with reported isolates. *Biochemical and Biophysical Research Communications* **226**, 314–318.

Heringlake, S., Osterkamp, S., Trautwein, C., Tillmann, H. L., Boker, K., Muerhoff, S., Mushahwar, I. K., Hunsmann, G. & Manns, M. P. (1996). Association between fulminant hepatic failure and a strain of GBV virus C. *Lancet* **348**, 1626–1629.

Jarvis, L. M., Watson, H. G., McOmish, F., Peutherer, J. F., Ludlam, C. A. & Simmonds, P. (1994). Frequent reinfection and reactivation of hepatitis C virus genotypes in multitransfused hemophiliacs. *Journal of Infectious Diseases* **170**, 1018–1022.

Jarvis, L. M., Davidson, F., Hanley, J. P., Yap, P. L., Ludlam, C. A. & Simmonds, P. (1996). Infection with hepatitis G virus among recipients of plasma products. *Lancet* **348**, 1352–1355.

Kao, J. H., Chen, P. J., Hsiang, S. C., Chen, W. & Chen, D. S. (1996). Phylogenetic analysis of GB virus C: comparison of isolates from Africa, North America, and Taiwan. *Journal of Infectious Diseases* **174**, 410–413.

Kumar, S., Tamura, K. & Nei, M. (1993). *MEGA: Molecular Evolutionary Genetics Analysis*, version 1.0. Pennsylvania: Pennsylvania State University.

Linnen, J., Wages, J., Zhangkeck, Z. Y., Fry, K. E., Krawczynski, K. Z., Alter, H., Koonin, E., Gallagher, M., Alter, M., Hadziyannis, S., Karayiannis, P., Fung, K., Nakatsuji, Y., Shih, J. W. K., Young, L., Piatak, M., Hoover, C., Fernandez, J., Chen, S., Zou, J. C., Morris, T., Hyams, K. C., Ismay, S., Lifson, J. D., Hess, G., Foung, S. K. H., Thomas, H., Bradley, D., Margolis, H. & Kim, J. P. (1996). Molecular cloning and disease association of hepatitis G virus: a transfusion-transmissible agent. *Science* **271**, 505–508.

McOmish, F., Yap, P. L., Dow, B. C., Follett, E. A. C., Seed, C., Keller, A. J., Cobain, T. J., Krusius, T., Kolho, E., Naukkarinen, R., Lin, C., Lai,

C., Leong, S., Medgyesi, G. A., Hejjas, M., Kiyokawa, H., Fukada, K., Cuypers, T., Saeed, A. A., Alrasheed, A. M., Lin, M. & Simmonds, P. (1994). Geographical distribution of hepatitis C virus genotypes in blood donors – an international collaborative survey. *Journal of Clinical Microbiology* **32**, 884–892.

Masuko, K., Mitsui, T., Iwano, K., Yamazaki, C., Okuda, K., Meguro, T., Murayama, N., Inoue, T., Tsuda, F., Okamoto, H., Miyakawa, Y. & Mayumi, M. (1996). Infection with hepatitis GB virus C in patients on maintenance hemodialysis. *New England Journal of Medicine* **334**, 1485–1490.

Moaven, L. D., Hyland, C. A., Young, I. F., Bowden, D. S., McCaw, R., Mison, L. & Locarnini, S. A. (1996). Prevalence of hepatitis G virus in Queensland blood donors. *Medical Journal of Australia* **165**, 369–371.

Muerhoff, A. S., Simons, J. N., Leary, T. P., Erker, J. C., Chalmers, M. L., Pilot-Matias, T. J., Dawson, G. J., Desai, S. M. & Mushahwar, I. K. (1996). Sequence heterogeneity within the 5′-terminal region of the hepatitis GB virus C genome and evidence for genotypes. *Journal of Hepatology* **25**, 379–384.

Ohba, K., Mizokami, M., Ohno, T., Suzuki, K., Orito, E., Ina, Y., Lau, J. Y. N. & Gojobori, T. (1995). Classification of hepatitis C virus into major types and subtypes based on molecular evolutionary analysis. *Virus Research* **36**, 201–214.

Pickering, J. M., Thomas, H. C. & Karayiannis, P. (1997). Genetic diversity between hepatitis G virus isolates: analysis of nucleotide variation in the NS-3 and putative 'core' peptide genes. *Journal of General Virology* **78**, 53–60.

Schlueter, V., Schmolke, S., Stark, K., Hess, G., Ofenlochhaehnle, B. & Engel, A. M. (1996). Reverse transcription–PCR detection of hepatitis G virus. *Journal of Clinical Microbiology* **34**, 2660–2664.

Schmidt, B., Korn, K. & Fleckenstein, B. (1996). Molecular evidence for transmission of hepatitis G virus by blood transfusion. *Lancet* **347**, 909.

Schreier, E., Hohne, M., Kunkel, U., Berg, T. & Hopf, U. (1996). Hepatitis GBV-C sequences in patients infected with HCV contaminated anti-D immunoglobulin and among iv drug users in Germany. *Journal of Hepatology* **25**, 385–389.

Shao, L., Shinzawa, H., Ishikawa, K., Zhang, X. H., Ishibashi, M., Misawa, H., Yamada, N., Togashi, H. & Takahashi, T. (1996). Sequence of hepatitis G virus genome isolated from a Japanese patient with non-A–E-hepatitis: amplification and cloning by long reverse transcription–PCR. *Biochemical and Biophysical Research Communications* **228**, 785–791.

Simmonds, P., Holmes, E. C., Cha, T. A., Chan, S.-W., McOmish, F., Irvine, B., Beall, E., Yap, P. L., Kolberg, J. & Urdea, M. S. (1993*a*). Classification of hepatitis C virus into six major genotypes and a series of subtypes by phylogenetic analysis of the NS-5 region. *Journal of General Virology* **74**, 2391–2399.

Simmonds, P., McOmish, F., Yap, P. L., Chan, S. W., Lin, C. K., Dusheiko, G., Saeed, A. A. & Holmes, E. C. (1993*b*). Sequence variability in the 5′ non coding region of hepatitis C virus: identification of a new virus type and restrictions on sequence diversity. *Journal of General Virology* **74**, 661–668.

Simmonds, P., Smith, D. B., McOmish, F., Yap, P. L., Kolberg, J., Urdea, M. S. & Holmes, E. C. (1994). Identification of genotypes of hepatitis C virus by sequence comparisons in the core, E1 and NS-5 regions. *Journal of General Virology* **75**, 1053–1061.

Simmonds, P., Mellor, J., Holmes, E. C., Tamprasert, S., Sakuldamrongpanich, T., Nuchaprayoon, C. & Smith, D. B. (1996). Evolutionary analysis of variants of hepatitis C virus found in South-East Asia: comparison with classifications based upon sequence similarity. *Journal of General Virology* **77**, 3013–3024.

Simons, J. N., Leary, T. P., Dawson, G. J., Pilot-Matias, T. J., Muerhoff, A. S., Schlauder, G. G., Desai, S. M. & Mushahwar, I. K. (1995). Isolation of novel virus-like sequences associated with human hepatitis. *Nature Medicine* **1**, 564–569.

Simons, J. N., Desai, S. M., Schultz, D. E., Lemon, S. M. & Mushahwar, I. K. (1996). Translation initiation in GB viruses A and C: evidence for internal ribosome entry and implications for genome organization. *Journal of Virology* **70**, 6126–6135.

Smith, D. B., Mellor, J., Jarvis, L. M., Davidson, F., Kolberg, J., Urdea, M., Yap, P. L., Simmonds, P. & The International HCV Collaborative Study Group (1995). Variation of the hepatitis C virus 5′ non-coding region: implications for secondary structure, virus detection and typing. *Journal of General Virology* **76**, 1749–1761.

Smith, D. B., Pathirana, S., Davidson, F., Lawlor, E., Power, J., Yap, P. L. & Simmonds, P. (1997). The origin of hepatitis C virus genotypes. *Journal of General Virology* **78**, 321–328.

Stark, K., Bienzle, U., Hess, G., Engel, A. M., Hegenscheid, B. & Schluter, W. (1996). Detection of the hepatitis G virus genome among injecting drug users, homosexual and bisexual men, and blood donors. *Journal of Infectious Diseases* **174**, 1320–1323.

Stuyver, L., Rossau, R., Wyseur, A., Duhamel, M., Vanderborght, B., Van Heuverswyn, H. & Maertens, G. (1993). Typing of hepatitis C virus isolates and characterization of new subtypes using a line probe assay. *Journal of General Virology* **74**, 1093–1102.

Tsuda, F., Hadiwandowo, S., Sawada, N., Fukuda, M., Tanaka, T., Okamoto, H., Miyakawa, Y. & Mayumi, M. (1996). Infection with GB virus C (GBV-C) in patients with chronic liver disease or on maintenance hemodialysis in Indonesia. *Journal of Medical Virology* **49**, 248–252.

Viazov, S., Riffelmann, M., Khoudyakov, Y., Fields, H., Varenholz, C. & Roggendorf, M. (1997). Genetic heterogeneity of hepatitis G virus isolates from different parts of the world. *Journal of General Virology* **78**, 577–581.

Received 24 January 1997; Accepted 25 March 1997